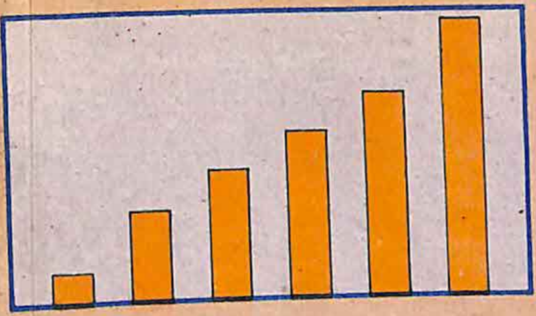
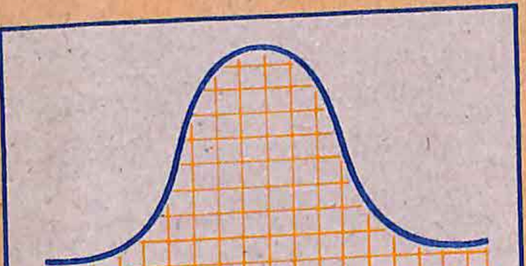


RAM PRAKASH GUPTA
PREM CHAND

Elementary STATISTICS



PITAMBAR PUBLISHING COMPANY

PREPARED IN ACCORDANCE WITH THE LATEST SYLLABUS OF CENTRAL BOARD OF SECONDARY EDUCATION, NEW DELHI FOR SENIOR SCHOOL CERTIFICATE EXAMINATION, FOR ALL INDIA & DELHI SCHEMES.

ELEMENTARY STATISTICS

(FOR CLASS XI)

By

RAM PARKASH GUPTA
M.A., B.T

&

PREM CHAND
M.A., M.Com., M.Ed., C.Ed.A. (B'ham)

3702

S.C.E.R.T. West Bengal
Library
Calcutta

PITAMBAR PUBLISHING COMPANY
Educational Publishers
888, East Park Road, Karol Bagh,
New Delhi-110005. (INDIA)

BY THE SAME AUTHORS

EVOLUTION OF INDIAN ECONOMY

Publishers
PITAMBAR PUBLISHING COMPANY
Educational Publishers
888, East Park Road, Karol Bagh,
New Delhi-110005 (INDIA)
Telephones : Off. 776058, 770067, 526933
Res. 5721321, 5715182, 586788

Edition 1989

Price Rs. 25/-

Code No. 27387

S.C.E.R.T., W.B. LIBRARY
Date 30.10.98
Accn. No. 9752

Copyright © Reserved with the Authors

Typesetter **JAYSHREE PRINTOGRAPH**

Printer
PIYUSH PRINTERS PUBLISHERS (P) Ltd.
G-12, Udyog Nagar,
Rohtak Road Industrial Area,
New Delhi-110041.
(Tel: 5472440)

PREFACE

"Elementary Statistics" covers Part II of the paper of Evolution of Indian Economy and Elementary Statistics meant for Senior Secondary students of Class XI as per the new CBSE syllabus. Part I entitled "Evolution of Indian Economy" has already gained popularity among students and teachers of economics within a month of its publication.

We hope this book will also be liked for its content, matter and format.

We have tried our level best to keep the subject matter within the limits of the time schedule assigned by the Central Board for the coverage of the subject.

This book is well illustrated with charts, diagrams and graphs. In the numerical part, each point has been explained and step-wise suitable solved examples have been incorporated in the text. At the end of each chapter graded questions are given so that the students may be able to practice whatever has been incorporated in the text.

We are confident that the present book will prove handy for the teachers teaching the subject and help the students to understand the various complexities of the subject. Suggestions for the improvement of the book from any quarter will be acknowledged with gratitude.

—Authors

SYLLABUS IN ELEMENTARY STATISTICS FOR CLASS XI PRESCRIBED BY CENTRAL BOARD OF SECONDARY EDUCATION, NEW DELHI

One Paper : *100 Marks 3 Hrs.

ELEMENTARY STATISTICS

Unit 4	Statistics in Economic Analysis	5
Unit 5	Collection and Organisation of Data	10
Unit 6	Presentation of Data	15
Unit 7	Condensation of Data	20
		50

Unit 4 : **Statistics in Economic Analysis**

Need for empirical and quantitative analysis.
Statistical tools used in economic analysis.

Unit 5 : **Collection and Organisation of Data**

Statistical data : primary and secondary data.
Methods of collecting data : census and sample methods
Reliability of sample data.
Organisation of data.

Unit 6 : **Presentation of Data**

Tables.
Diagrams : Bar diagrams, pie diagrams, frequency curves and ogives.
Time series.

Unit 7 : **Condensation of Data**

Measures of central tendency : mean, mode and median.
Partition values : quartiles, deciles and percentiles.
Measures of variability : range, mean, deviation, standard deviation and coefficient of variation.
Measures of Correlation (two variables and ungrouped data only)
Index numbers.

Note: *50 marks are allotted to Units 1,2 and 3 covering Evolution of Indian Economy.

CONTENTS

Chapter Page

UNIT I—Statistics in Economic Analysis

1. Need for Empirical and Quantitative Analysis 1
2. Statistical Tools used in Economic Analysis 8

UNIT II—Collection & Organisation of Data

1. Statistical Data : Primary and Secondary Data 20
2. Methods of Collecting Data 46
3. Organisation of Data 62

UNIT III—Presentation of Data

1. Presentation of Data-Tables and Diagrams 83
2. Presentation of Data-Time Series and Graphs 111

UNIT IV—Condensation of Data

1. Measures of Central Tendency—Averages 137
2. Measures of Central Tendency—Median 158
3. Measures of Central Tendency—Mode 173
4. Partition Values 185
5. Measures of Variability 197
6. Measures of Correlation (Two Variables & Ungrouped Data) 237
7. Index Numbers 257

Appendix

UNIT I

STATISTICS IN ECONOMIC ANALYSIS

1. Need for empirical and quantitative analysis
2. Statistical tools used in economic analysis

1

Need For Empirical And Quantitative Analysis

Statistics has been in constant use since long. Initially, it was concerned with State Administration. The kings or monarchs used it for the collection and analysis of data concerning population, birth and death rates, military power or force, agricultural land, state wealth etc. Since the use of statistics was limited only to the governments, it was called *"State Science" or "Administrative Art ". In modern times its use is no longer confined to governments. Rather it is increasingly used in Physical Science, Life Science and Social Sciences. In fact, statistics is being used freely in all aspects of human and social life along with all branches of science. Statistics is also increasingly used in economics in recent years. The statistical measures such as measures of central tendency, dispersion, skewness*, and *correlation* etc. are widely used in different branches of human activity.

ECONOMIC ANALYSIS

Economics is a science which studies human behaviour in relation to scarce means having alternative uses. While studying human behaviour in relation to economic situation we are to keep in mind various economic variables such as prices and demand for commodities, family income and expenditure and price level and cost of living etc. This is possible through statistical analysis alone. Statistical analysis provides credibility to economic analysis. Through statistical analysis alone the economists test the relationship between various economic variables. Economic analysis becomes effective with the help of statistical analysis.

Forms of Analysis

Analysis is divided into '*Empirical Analysis*' and 'Quantitative Analysis'. They are briefly explained hereunder:-

(a) *Empirical Analysis*: It is a method under which a subject is studied on the basis of observations or experiments. Under natural sciences such as physics, chemistry and biology this method is commonly used and relevant laws and statements can be tested and experimented in laboratories. As against this, economics is a social science wherein human behaviour is the subject of study and it covers the whole world. Empirical analysis is not helpful for the study of economics and other social sciences. The problems of social sciences can be effectively studied through the technique of quantitative analysis. Under quantitative analysis technique the problem can be studied systematically, but the results so obtained cannot be tested in a laboratory and hence they are not so exact and accurate as the results of empirical analysis

(b) *Quantitative Analysis*: An effort is made to provide preciseness to the facts through quantitative analysis. Under it we study facts in the form of quantities such as 25, 30, 35 as numbers, in weight we use kilograms, quintles and tons etc., and in height centimetre, metre etc. are used. The numerical data is then classified, tabulated and analysed so that reasonable conclusions be drawn. In short, the main purpose of '*quantitative analysis*' is to provide preciseness to the facts so that they can be easily compared. Suppose we want to study the relationship between cost and supply of a particular commodity. It will be possible only when sufficient data regarding cost and supply of that particular commodity is collected, tabulated, classified and analysed.

MEANING OF STATISTICS

The word statistics is commonly used in two ways - singular and plural. In the plural sense it stands for numerical or quantitative data such as the number of unemployed persons in various states of India, details of production in various crops, details of illiterate persons in various states of India and so on. Hence, statistics in the plural sense stands for *numerical data* only. However, when statistics is used in the singular form, it refers to the "*Science of Statistics*". The science of statistics helps us in the collection of necessary data, in its presentation and analysis. It also throws light on the principles and methods used in the process of collection, presentation and analysis of data.

Definition of Statistics

Various authorities have defined statistics in different ways. Some have defined it in the singular form while others have explained and defined the subject in the plural form.

Statistics in the Singular Form

Dr. A.L.Bowley has defined statistics in various words and ways. At one place he defined statistics as a *"science of counting"*. At another place he pointed out and stated that *"statistics is a science of averages"*. Both these interpretations are very limited because statistics is neither the science of counting alone nor the science of averages. Dr. Bowley has defined statistics at another place as *"numerical facts in any department of enquiry placed in relation to each other"*. In this definition also only some aspects are highlighted and not all. Another definition of statistics given by Dr. Bowley is: *"Statistics is the science of measurement of social organism regarded as a whole in all its manifestation"*.

The above definition of Dr. Bowley too is not universally acceptable because it has limited the scope of statistics to social sciences alone. Even Dr. Bowley himself agrees with the view when he says that science of statistics cannot be limited to any one particular science or subject of study. Professor Bodington has defined statistics as: *"Statistics is the science of estimates and probabilities"*. This too is a very narrow and limited definition. Croxton and Cowden have defined statistics in simple words against a big canvass. According to them *"Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data"*. Professor Saligman has defined statistics still in a bigger context. According to him *"Statistics is the science which deals with methods of collecting, classifying, presenting, comparing and interpret ing numerical data collected to throw some light on any sphere of enquiry". The definition given by Croxton and Cowden is quite comprehensive as it specifies the different stages through which a statistical inquiry must pass.*

Statistics in the Plural Form

Webster has defined statistics in the pluraal sense of the term. According to him *"Statistics are the classifed facts representing the conditions of the people in a state especially those facts which can be stated in number or in tables of numbers or in any other tabular or classified arrangement"*. This definition is not a complete definition of statistics. It limits the scope of statistics only to the use of data collected by the various state authorites for use in administraion. In reality the scope of statistics has widened considerably.

Professor *Horace Secrist* too has defined statistics in the plural sense. According to him *"by statistics we mean aggregate of facts affected* *to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".*

In the words of Conner, Statistics are *measurements, enumerations or estimates of social or natural phenomena systematically arranged so as to exhibit inter-relationship* .

CHARACTERISTICS OF DATA (STATISTICS)

The definitions given above clearly point out certain characteristics which numerical data must possess so that they may be called statistics. These are as follows :

(a) *Numerically Expressed:* All statistics are numerical statements of facts i.e. expressed in numbers. The statements such as 'the population of India is rapidly increasing' or 'the production of maize is not sufficient' or 'India is a poor country', do not constitute statistics. The reason is that such statements are vague and one cannot make out anything from them. On the other hand, the statements like the production of foodgrains was 144.1 milllion tones in 1986-87, the population of India in 1984 was 758.16 million people are statistical statements.

(b) *Affected by Multiplicity of Causes:* Generally speaking facts and figures are affected to a considerable extent by a number of forces operating together. If we say that the production of sugar in 1986 is 85 lakh tonnes and in 1951, it was 36 lakh tonnes, then we will have to assume that the production is affected not only by one reason but by a number of reasons. Statistics of production of sugar are affected by rainfall, quality of soil, seeds and manure, methods of cultivation etc. It is very difficult to study separately the effect of each of these forces on the production of sugar. If we say that the average life expectancy of an Indian is 54 years and the average life expectancy of an inhabitant of Norway is 73 years, it is due to a number of reasons. Average life expectancy is affected by per capita income, standard of living, saritation etc.

(c) *Aggregate of Facts:* Single and isolated figures are unrelated and cannot be compared. To illustrate, a single figure related to a road accident cannot be regarded statistics because of its non-comparability and lack of relationship as parts of a common phenomenon.

(d) *Enumerated or Estimated:* Facts and figures about any phenomenon can be obtained in two ways, namely by actual counting or by estimate. If the area of our inquiry is small, then we will collect the facts and figures by the method of actual counting. On the other hand, if the area of inquiry is large, then we will collect the facts and figures by the method of estimate. Estimates cannot be as precise and accurate as actual counts or measurements.

(e) *Reasonable Standards of Accuracy:* It is clear that mathematical accuracy is not possible in collection or estimation of statistics. But reasonable standard will be different for different inquiries, viz. if we are measuring the heights of a group of students, data should be collected to the nearest centimetre but if we are estimating the distance between Bombay and Calcutta, accuracy upto nearest centimetre is not desired. In such a case, it may be true to the nearest kilometre.

(f) *Pre-determined Purpose:* The purpose of collecting data must be decided in advance. The purpose should be well defined and specific. A general statement of purpose is not enough. If the purpose is not pre-determined then a number of undesireable items will be included in and a lot of money will be spent on the survey. For example, if the objective is to collect data of the wages of industrial workers, it would not serve any useful purpose, unless we know for what purpose we want to collect data, whether we want to compare the wages of different workers, or we want to compare the standard of living of the workers etc. Thus, purpose of collecting data should be well defined.

(g) *Systematically Arranged:* Before collecting statistics a suitable plan of data collection should be prepared and the work carried out in a systematic manner. Data collected in a haphazard manner, would very likely lead to fallacious conclusions.

(h) *Placed in Relation to Each other*: If numerical facts are to be called statistics, they should be comparable. Valid comparisons can be made only if the data are homogeneous i.e. they relate to the same phenomenon or subject. It would be meaningless to compare the height of horses with the height of human beings. Similarly, it would be meaningless to compare the production of Premier Padmini Cars with the production of foodgrains.

STAGES OF STATISTICAL STUDY

There are five stages of statistical study. *First stage* is *collection of data* which refers to gathering statistical facts by different methods. *Second stage* is *to organise* the data so that collected information is easily understood. By organisation of data we mean arrangement of data in a systematic order. *Third stage* is *presentation of data*. After collection and organisation ,the data is to be reproduced by various methods of presentation so that different characteristics of data can easily be understood on the basis of their quality and uniformity. *Fourth stage* is the *analysis of data*. Different methods and tools are used for analysing statistical data. On the basis of these measures we try to interpret the results of the inquiry.

Thus, statistics is the scientific methodology of collection, organisation, analysis and interpretation of numerical data.

NEED FOR EMPIRICAL AND QUANTITATIVE ANALYSIS

Empirical analysis may be defined as a *"METHOD of studying a subject whereby knowledge is acquired as a result of actual experience."* (Hanson's dictionary for Economics and Commerce). In other words an empirical investigation is a study of a phenomenon through observation of facts. It is based on one's subjective impressions. All the natural sciences like physics and chemistry make use of the method. The scientists observe natural phenomenon and arrive at generalised conclusions. This method is being incresingly made use of in social sciences as well. Since the conclusions are based on experience rather than on mere subjective impressions, they have greater validity and acceptability. Take for instance the reaction of a person to weather. He says *"Today is colder than yesterday"*. It is a subjective impression. We have to observe the temperature on these two days and then arrive at a decision as to whether today is colder than yesterday.

An important function of statistics is to make the facts precise. In order to do this, the facts should be expressed in figures. i.e. in quantitative terms and not in qualitative terms. Let us consider for example a statement, "The production of wheat in 1989 is expected to be more than that of 1983." This statement gives only a partial picture of the Production situation. On the other hand if we quantify the statement as *"The production of wheat in India is expected to increase by 25% in 1989 over 1983"*, it conveys more definite information. Similarly statements like *"there is no dearth of educated unemployed in India"* or *"the population of India is growing rapidly"* or *"The prices of various commodities are rising"*, hardly carry any worthwhile information and thus make little sense. When the information is supplied in quantitative terms rather than qualitative we can be very precise and definite. As a result of which comparison becomes possible and also helps us to establish inter-relationship with the result already achieved.

In view of the importance of quantitative analysis Lord Kelvin had stated: *"What you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers your knowledge is of a meagre and unstatisfactory kind."*

DISTRUST AND MISUSE OF STATISTICS

We have seen above that statistical data and statistical techniques have come to stay as most important tools of mankind in its social, political, economic and business environment. In fact, Prof. Tippett was tempted to remark, *"Our very entry into the world and exit from it are* *recorded statistically." But who could reconcile this statement with the opinion of those who have developed an eyesore in regard to the use of statistics. They have always looked upon statistics with doubt and suspicion. Look to the specimen of misunderstood remarks against statistics:*

i) 'There are three kinds of lies-lies, damned lies, and statistics.'

ii) 'Statistics is like clay of which one can make a God or Devil, as one likes.'

iii) 'Statistics can prove or disprove anything.'

iv) 'Statistics are like biknis. They reveal what is interesting and conceal what is vital.'

Likewise, Stephen Leacock has remarked, *"In earlier times people had no statistics and so they had to fall back on lies. Hence the huge exaggerations of primitive literature giants or miracles or wonders. They did it with lies and we do it with statistics, but it is all the same."* A similar type of remark has been made by Derell Huff in his book, *"How to lie with statistics" when he writes, "A Well-wrapped statistics is better than Hilter's 'Big-Lie' it misleads, yet it cannot be pinned on you."*

The distrust in statistics arises because of its misuse. Professor Ya-lun-Chou rightly puts it as, *"Derogatory statements have been inspired by misuse of statistics on the part of untrained practitioners of statistical methods, ignorant users of statistical results, dishonest peopole who employ statistics to mislead and confuse rather than to clarify and illuminate."*

QUESTIONS

1. "Statistics is the art of comparison." Explain this statement.

2 Briefly define statistics.

3. What is the subject matter of Statistics ?

4. What is the importance of statistics in modern economic set-up?

5. Distinguish between:

a) 'Statistics' in the singular sense and in the plural sense.

b) External and internal sources of data.

6. Illustrate the need for empirical analysis in economics with two examples.

7. "Statistical analysis is of vital importance for successful businessmen, economists, administrators and educationists". Discuss with illustrations.

8. "Statistical thinking is as necessary for efficient citizenship as the ability to read and write." Explain this statement in about 200 words.

2

Statistical Tools Used In Economic Analysis

Statistics is a science which has necessary tools for analysis. If economic analysis is to be carried out through statistical methods, following are some of the important statistical techniques, which are applied in economic analysis:

STATISTICAL TOOLS

(a)	(b)	(c)	(d)
Collection of data, Census or Sampling *Source* Primary and Secondary	Organisation of data, Editing, Classifying and Tabulation.	Presentation of data : Tables, Graphs, Diagrams.	Analysis of data; Central tendency, Dispersion, Correlation and regression. Time series, Index Numbers, Interpolation and extrapolation.

(a) Collection of Data: This is the first stage in any statistical investigation. By collection of data, we understand the method employed in the collection of data. In collecting the data, the census or sampling methods are employed. These methods are commonly used in economic analysis; data relating to the population of any country is collected by census method, whereas the National Income is estimated by sampling method. In market surveys, research and other investigations, the sampling methods are generally employed. The economic data can be collected by primary or secondary sources. Great care is to be taken to collect data as they constitute the basis for other stages. Faulty data will lead to defective results.

(b) Organisation of Data: The second stage involves editing, classification and organisation of the collected data. Quite often, the

statistical data collected is in the form of schedules or questionnaires, which contain a mass of raw material and details, without any form or structure, it requires organisation in a form that is easily comprehensible. The best method of organising the data is to group this data suitably and arrange it in some sort of condensed form. The first step in the organisation of data consists in classifying the data according to some common characteristics of the variable. This process of grouping into different classes or sub-classes according to the different characteristics of the variable is called classification. When the organised or classified information is presented in precise and systematic tables, it is called tabulation. The process of classification, arranging data in a certain logical and/or in statistical series and presenting them in tabular form is called the organisation of data.

(c) Presentation of Data: The statistical techniques of presentation of data like tables, charts, diagrams and graphs help in economic analysis. Almost all the economic concepts are explained in tabular, diagrammatic and graphic manner. The basic concept of economics and the Law of Demand cannot be explained without the use of tabular and graphic method. The concept of price determination is based on graphic presentation. The targets and the productivity in various sectors is generally presented by tabular, diagrammatic and graphic methods.

(d) Analysis of Data: There are various methods of analysis of data, each has its definite place in economic analysis. Following are some of the methods, which are used in economic analysis:

(i) Use of Measures of Central Tendency in Economic Analysis : Central tendencies which are popularly known as averages, have a unique place in economic analysis. There are various types of averages, the most popular ones are arithmetic mean, harmonic mean, geometric mean, median and mode. Each has got its own significance in economic analysis e.g. if we want to know the average income, average consumption, average production, etc. the arithmetic mean is used. If the things are expressed in ratio, for example, 10 units per rupee and 16 units per rupee etc. or where the prices are different and the amount spent on them in each case is the same, the harmonic mean is to be used. For example, if a man purchases two types of apples at Rs.10/-, in this case, harmonic mean will give us the correct average. If we want to measure the growth rate of population, the geometric mean is to be used. Whenever, the modal life of a plant, machinery etc. is to be ascertained, the mode is used. '*A fashionable item' 'most popular items*' always refer to mode. In measuring the average intelligence, average height, average weight or other qualitative phenomenon median is to be used. We shall discuss, in detail, the applicability of each measure of Central Tendency in subsequent

chapters.

(ii) Use of Measure of Dispersion in Economic Analysis : Central tendency throws light on the central position of the series. Two different series may have the same means but they may differ widely as to their concentration. The meausres of dispersion are used in knowing the wideness of individual item from the central tendency. The disparities or inequalities of income distribution are calculated with the help of this statistical technique.

Skewness measures lack of symmetry in the distribution of items. It could be used for knowing the concentration of maximum items, whether most of them are associated with smaller items or larger items. This technique is useful in knowing the concentration of wealth, whether larger number of people have smaller amount of income or smaller number of people have got the maximum amount of national income.

Kurtosis measures the concentration of frequencies near the central value. If the large number of popele have the average income, the curve drawn will be known as 'Leptokurtic', if a smaller number of people have average income the curve drawn will be 'mesokurtic'. Thus, it measures the concentration of items around or at the central value.

(iii) Use of correlation and regression in Economic Analysis : The relationship between two or more variables can be known by correlation and regression techniques of statistics. The relationship between input (of a variable) and the output can be studied with the help of correlation between the two variables. It may indicate that an increase in one is associated with the decrease in another. Here the variables are inversely related. For example, price-demand relationship is negative. A positive relationship exists between income and consumption which indicates that with the increase in income consumption also increases.

Once we have established the nature of relationship between two variables, we can predict the behaviour of one variable in terms of another by the technique of regression analysis.

Regression analysis serves as a tool in forecasting the nature of a phenomenon.

(iv) Use of Time Series in Economic Analysis: The time series analysis aids in the important problem of forecasting. It is of great importance in economic analysis. It helps in studying the fluctuations in a period of number of years. Secular variations reflect the long-range growth or decline of a phenomenon. It helps in determining the policy of an economy in various spheres. Seasonal variations help businessmen to foresee the sale of a particular product during a year. It helps them in adapting themselves to the changed conditions.

(v) Use of Index Numbers in Economic Analysis: Index numbers are generally called as 'Economic Barometer'. As the weather barometer describes the weather conditions, similarly index numbers describe the economic and business conditions. There are various types of index numbers like price index numbers which describe the changes in price level over a period of time, the cost of living index numbers which show whether real wages are rising or falling, consumer price index numbers are useful, particularly, to employers, employees, trade unions, and the Government to decide the dearness allowances. Similarly, foreign trade index numbers, industrial and agricultural production index numbers have got special place in economic policies of the government.

(vi) Use of sampling in Economic Analysis: When the complete investigation of a population becomes impossible, the sampling technique is used. There are certain instances, where the census method of investigation becomes impossible or cannot be properly used. In those cases sampling is generally used. For example, we have to test the sharpness of blades and if we test each blade, perhaps whole of the product will be wasted. In such circumstances, the census method will not be suitable. Under these circumstances, sampling techniques will be more useful. The sampling techniques are being widely utilised now-a-days. A cart load of coal is accepted or rejected on the evidence gained from testing only a few units. The doctor makes inferences about the patient's blood through the examination of a few drops of blood. Similarly, in the study of different economic variables, like agricultural productivity, industrial output, size of the firms etc. sampling may be found more useful. Thus, the sampling technique is unique in ascertaining the full knowledge of the universe with the help of a few observations.

(vii) Use of Interpolation and Extrapolation in Economic Analysis: Interpolation is a device by which, we can measure the probable value of given phenomenon in between the two or more figures. For example, census in India is taken after a decade and if we want inter-census population data, the technique of interpolation is used. The insurance companies fix the rate of premium with the help of interpolation technique.

Extrapolation offers a device for predicting the future probable value of the phenomenon. The probable sales at different prices will be ascertained by extrapolation device.

DIVISIONS OF STATISTICS

(i) Theoretical Statistics: In theoretical statistics all those principles are studied on which statistics is based. All the statistical methods are studied therein.

(ii) Applied Statistics: From statistical methods, we get

knowledge about principles. Here, we study how to make use of these principles in practical life. In other words, in applied statistics, we study how to make proper use of all the statistical methods. In the words of George Simpson and Fritz Kalfka *"The applied statistician puts statistical methods into practice in a particular subject matter."*

iii) Descriptive Statistics: Statistical measures such as those of averages, dispersion, and skewness etc. are called descriptive statistics since they are helpful in understanding the characteristic features of a given statistical series.

LIMITATIONS OF STATISTICS

Though statistical methods are now extensively used almost in all the fields of human inquiry, yet these methods have their own limitations. If these limitations are not taken into consideration, the results will be doubtful. Statistics must be regarded as an instrument of research of great value but having several limitations which are not possible to overcome and as such, these limitations need our careful attention. The main limitations of statistics are as under:

1. Statistics does not study Individuals: Statistics deals only with aggregates. To quote W.I.King here, *'Statistics largely deals with averages and these averages may be made up of individual items radically different from each other*". No significance is attached to individual items which constitute the aggregate. Suppose five persons live at a particular place. They are A, B, C, D, & E, and their monthly incomes are Rs.400/-, Rs.600/-, Rs.1000/-, Rs.1200/-, and Rs.300/- respectively. The average income comes to Rs.700/-. We come to the conclusion that the people of that place are happy and prosperous. The economic condition of person 'E is very poor. But we do not give importance to this. Thus, statistical conclusions explain the aggregative behaviour of the group and throw no light on the individual behaviour of the constituents of the group.

2. Statistical laws are not exact : Unlike the laws of physical and natural sciences, statistical laws are only approximations and not exact. On the basis of statistical analysis we can talk only in terms of probability and chance and not in terms of certainty.

3. Statistics deals with quantitative characteristics: Statistics are numerical statements of facts. Such characteristics which cannot be expressed in numbers are incapable of statistical analysis. Thus qualitative characteristics like honesty, efficiency etc. cannot be studied directly. However, it may be noted that attempts have been made to measure these phenomena indirectly. In the words of Bowley, *"Statistics is the science of measurements of social organism, regarded as a whole in all its manifestations."* We study only the quantitativeaspect of the problem and not the qualitative aspect of the problem.

4. Statistical results are true only on an average: The conclusions obtained statistically are not universally true. They are true only under certain conditions. This is because statistics as a science is less exact as compared to natural sciences.

5. Statistical conclusion can be mis-interpreted: Statistics can be used to establish conclusions which suit the investigators. As the saying goes, *"Statistical methods are the most dangerous tools in the hands of inexperts."* The use of statistical tools by inexperienced and untrained persons might lead to very fallacious conclusion. As King says, *"statistics are like clay of which one can make a God or Devil as one pleases."*

6. Only one set of method of studying a problem: Statistical methods constitue only one set of method of studying a problem. There are other methods also. These methods should be used to supplement the conclusions arrived at with the help of statistics. Statistical methods may in some cases fail to reveal the entire story. Many problems are affected by such factors which are incapable of statistical analysis and so it is not possible to examine a problem in all its manifestations only with the help of statistical approach.

7. Statistics cannot prove anything: Some statisticians claim that statistics cannot prove anything. In fact, they only describe a phenomenon quantitatively, classify it into parts, summarise a fact relating to each part and prepare the ground for logical inference. But the inference must be drawn in the appropriate surroundings which may be personal, political or social, otherwise it would not be valid.

FUNCTIONS OF STATISTICS

In modern times Statistics is viewed not as a mere device for collecting numerical data but as means of developing sound techniques for their handling and analysis and drawing valid inferences from them. To draw valid conclusions is the function of a Statistician. *Robert W.Burgess* has beautifully summed up the functions of statistics as, *"The fundamental gospel of statistics is to push back the domain of ignorance, prejudice, rule of thumb, arbitrary or premature decisions, traditions and dogmatism, and to increase the domain in which decisions are made and priciples are formulated on the basis of analysed quantitative facts."* The following are the important functions of statistics:

1.It presents facts in a definite form: Numerical conclusions are convincing and definite. For example, if we say that America is a very rich country, it is not a definite conclusion. If we say that the per capita income of America is 10 times more than the per captia income in India, it is a definite conclusion because it shows difference in income between two countries.

2. *Statistics simplifies complexities*: Statistical techniques such as tabulation, diagrammatic and graphic presentation of data, averages, percentages, index numbers etc. make the facts and figures easily understandable. Data concerning crores of people can be presented in a few classes. This solves the problem of complexity.

3. *It facilitates comparison:* Unless figures are compared with others of the same kind, they are often devoid of any meaning. For example, to state that the population of India in 1986 is 75.8 crores hardly means anything unless the figure is compared with earlier time periods or with population of other countries.

4. *It helps in the formulation of policies* : Statistics renders a useful service to the decision making authorities by providing the necessary statistical information for framing suitable policies. It helps in the formulation of policies in all the fields i.e. economic, social and trade etc. The export and import policy, the price policy, wage policy and other policies depend upon statistics.

5. *Statistics enlarges individual experience and knowledge*: Statistics is a body of methods which widens the sphere of human knowledge and which adds precision and exactness to our knowledge. With the help of statistics man develops his ability and efficiency. Statistics gives clarity and definiteness of certain concepts and views. Bowley remarks- *the proper function of statistics is indeed to enlarge individual experience*.

6. *Statistics tests the laws of other sciences:* Statistical methods are extremely helpful in formulating and testing hypothesis and to develop new theories on the basis of old theories and new laws are formed. Certain changes have been made in the Malthusian Theory of population and Quantity Theory of Money on the basis of new data.

7. *Statistics helps in forecasting future events* :Statistics helps in forecasting the future events by estimating the relationship between variables that existed in the past. There are special statistical techniques such as extrapolation and time-series analysis which enable us to peep into the future.

8. *Statistics explains relationship between two or more variables:* An important function of statistics is to investigate the relationship between two or more variables. In the absence of statistical methods it would be very difficult to draw an accurate and wise conclusion. Bowley is of the opinion, "*A chief practical use of statistics is to show relative importance, the very thing which an individual is likely to misjudge, statistics are always comparative.*"

IMPORTANCE OF STATISTICS

The science of statistics has a wide application and has become indispensable in various fields of inquiry. The knowledge of statistics has become a must. We must learn statistics in order to understand our daily problems and the world around us. In the words of A.G.Wells, "*Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.*" That day has already come. In the ancient period, government employees had to collect the data to tell the government about its working. At that time its use was limited to state only. Now it is not confined to the affairs of the state but is intruding constantly into various diversified spheres of life-social, economic and political. It is now finding wide applications in almost all sciences-social as well as physical such as biology, psychology, education, economics, business mangement, astronomy, mathematics etc. It is hardly possible to enumerate even a single department of human activity where statistics does not creep in. It is also called the Arithmetic of Human Welfare. In the words of Croxton and Cowden, "*We do not find any aspect of human activities where statistics is not used.*"

The importance of statistics can be explained with the help of *following points:-*

a) Statistics is arithmetic of human welfare: Statistics is indispendable for a clear appreciation of any problem affecting the welfare of mankind. Problems relating to poverty, unemployment, food shortage etc. cannot be fully weighed without the statistical balance. In the words of Bowley, *"Statistics is the tool in the hands of mankind to translate a multitude of complex facts into simple and understandable statement of facts."*

b) Essential for Planning: It is an era of planning. Every country, rich or poor is engaged in planning. It is very difficult to prepare a plan without the correct data. Planning is very much based on assumed figures. From the figures of income and expenditure we come to know about the priorities. Our targets are also based on those statistics which are indispensable in planning and the various decisions of the Government with regard to import, export, production etc. are all based on facts and figures. In fact statistics serves as foundations of superstructure of planning.

c) Helpful in Administration: There was a time when State used to collect statistics only for measuring military strength and wealth of the nation. But there has been manifold increases in the functions of the State these days. Statistics is very helpful to a state as it helps it in administration. Modern state makes extensive use of statistical data on various problems. Before enforcing any policy, State has to examine its merits and demerits and this can be done only with the help of numerical data. The Finance Minister prepares the budget by having the data regarding income and expenditure. He imposes more taxes on the rich people. The State is always the most important single unit which not only collects the largest amount of statistics but also needs statistics on a very extensive scale. Statistics are so significant to the state that the government in most countries is the biggest collector and user of statistical data. Such data is of immense help to many institutions and research scholars who further process it and arrive at useful conclusions which help in decision making.

d) Helpful in Business and Commerce: Some knowledge of statistics is essential today for business. Statistics helps the businessman in the formulation of policies with regard to business. Production and sales policies can profitably be planned with the help of statistics and failures due to wrong expectations can be avoided. For planning of production, scale of operation, quality control, plant layout and location, for study of changes in fashions and tastes, allocation of advertisement funds and for study of various business trends, there is nothing more useful than Statistics. A decision on location will depend on the collection and analysis of information on the availability of cost of land, raw-material etc. A decision on size is also necessary because cost is immediately connected with the scale of operations. The techniques of time series analysis and business forecasting enable the businessman to predict with a fair degree of accuracy the effect of a large number of variables. In fact Statistics is highly useful in business that a prominent business executive and Statistician said thirty years ago *"When the history of modern times is finally written, we shall read it as beginning with the age of steam and then progressing through the age of electricity to that of statistics."*

e) Utility to Social Workers: Statisticss is of great utility to the social reformers. They study the problems and analyse them on the basis of data. For example, they study how many deaths have taken place due to dowry, how many industrial workers drink, how many farmers are under the clutches of the landlords or money-lenders. They tell the government about these evils and suggest measures. The laws are framed on these basis. Thus statistics helps us in framing laws for eradicating social evils.

f) Tool of Management: Statistics helps the businessman in the management of the business. The success of production programming both in the short as well as in the long period depends to a great extent on the quality of sales forecasts and projections. The variations in sales below or above the forecast may disturb the production schedules. Effective control on sales can be exercised through regional allocations. Statistical methods also come to the aid of quality control engineer. The knowledge of Statistics will help the businessman in adopting a suitable policy about the exact increase or decrease of the stock of a commodity. Statistical techniques can serve as the best guide for solution of other business problems.

g) Utility of Statistics to Research Workers: Statistical analysis is one of the most powerful tools in the hands of the modern research workers. Experiments about crop yields with different types of fertilisers and different types of soils are very often designed and anlysed according to statistical methods. In the field of industry and commerce, statisticians carry on different types of researches. They try to find out the sources and causes of variations of different products from their standard quality.

h) Utility of Statistics to Bankers, Brokers and Insurance companies: The business of banking and insurance is not possible without the use of statistical analysis because it is needed for fixing money rates, establishment of new branches, grant of loans, fixation of the premium rates etc. Insurance companies cannot carry on their business in the absence of statistical data relating to life and mortality tables and premium rates etc. Statistics also helps the bankers in deciding how much cash they should keep in their vaults to meet day to day demands of the customers.

i) Helpful to understand Economic Growth: Statistics helps us in finding out the rate of growth of the economy. We come to know from the figures about the rate of growth of national income, rate of growth of agriculture, and industry , etc. These figures show whether the country is developing or not developing. We also make comparisons on the basis of statistical information.

j) Statistics has universal utility: In the modern age statistics has great utility. The knowledge of statistics is very useful for the development of common man. Statistics is indispensable in Economics. It is almost impossible to find a problem in Economics which does not require an extensive use of statistical data. Statistics is the tool in the hands of mankind to translate a multitude of complex facts into simple and understandable statement of facts. Without statistics the facts may remain remote from the domain of our understanding. Statistics lends precision to our ideas. It enriches and enlarges individual experience. Statistics increases the field of mental vision of an opera glass or telescope and also increases the field of physical vision.

STATISTICS AND ECONOMICS

Statistics is very useful in the study of Economics. Many economists have used statistical tools for economic research. In India, Prof. P.C.Mahalanobis, Dr. V.K. R.V. Rao, R.C. Desai etc. have extensively used statistical tools for day to day economic analysis. There is a vital relationship among economics, mathematics and statistics due to the importance of mathematics and statistics in economics. As a result, new science is evolved which is called Econometrics.

Many important laws of economics have been evolved due to statistical analysis e.g. Engel's law of consumption, Malthusian Theory of population etc. The reliability of economic laws is being tested by using statistical tools. Statistical methods have made a contribution to the development of empirical side of economics; the inductive method of studying economics is wholly dependent upon statistical methods.

Like a doctor of medicines, the economist also uses statistical tools to examine various economic data. It becomes easier to understand various economic problems by the use of statistical tools. It helps in formulation of economic policies. Professor Marshall has said, *"Statistics are the straw art of which I, like every other economists, have to make bricks".* Similarly Tugwell said, *"The science of economics is becoming statistical in its methods."*

Traditional economists have divided the study of economics into five parts viz. production, consumption, exchange, distribution and public finance. Let us understand the importance of statistics in the study of various parts of economics.

i) Statistics and Production: With the help of statistics, we can study the productivity of various factors of production. The progress of production can be easily measured by statistics. The statistics of production are very helpful for adjustment of demand and supply. In every developed country, the census of production is taken with a view to make a comparative study of various fields of production and economic planning.

ii) Statistics and consumption: With the help of statistics, we can find out how different groups spend their income on consumption. Statistical data has helped us to know that individuals spend their income in a particular order i.e. first on necessities, then on comforts and later on luxuries according to the income of the individuals and their standard of living.

iii) Statistics and Exchange: A producer needs statistics on the cost of production for deciding selling price. He needs data to study competition and demand for his product. The law of price determination and cost price which are based on the market conditions and demand and supply can be studied with the help of statistics.

iv) Statistics and Distribution :Statistical methods are used in solving the problem of the distribution of the national income. We know a number of problems arise due to unequal distribution of wealth and national income. They can be solved with the help of statistical data.

v) Statistics and Economic Planning: These days almost all the countries of the world adopt the method of planning for economic development. Statistics is the most important tool in economic planning Statistics helps economic planning by collecting data of national resources both human and natural. For drawing a plan of their use, we need the help of statistics. It is also needed at the time of execution of the plan and also at the time of analysing the progress of the plan. With the help of statistical data, we can compare the stage of economic development of one country with the other. Various problems of under-developed countries like problem of over population, lack of industries, lack of agricultural development and lack of education can be studied by getting figures with the help of statistics. We can view the progress of a country in various sectors with the help of statistical data and statistical methods. Priorities of expenditure of a national budget can be determined through the comparative study of past performances of various ministries and departments with the present. Thus, planning needs help of statistics at every stage.

Besides the above fields, statistical tools are indispensable for many other fields such as Sociology, Mathematics etc. Bowley has rightly said, *"A knowledge of statistics is like a knowledge of foreign language or algebra, it may prove useful at any time under any circumstances."* The fact that in the modern world statistical methods are of universal applicability is in itself enough to show how importatant the science of statistics is. In the words of Croxton and Cowden, *"There is no field of life where statistics is not used."*

QUESTIONS

1. Describe the importance of statistics in economic planning. What are its limitations ?

2. Describe in brief the important statistical tools used in economic analysis.

3.. Explain the important functions of statistics with suitable illustrations.

4. Distinguish between:

a) Interpolation and Extrapolation.

b) Collection and presentation of data.

S.C.E.R.T. W.B. LIBRARY
Date ....
Accn. No..........

5. What are the important functions of Statistics ?

6. In what way Statistics is useful to economists, the industrialists and traders.

7. "There is no field of life where statistics is not used." Explain this statement in about 200 words.

8. Discuss the role that statistics can play in the process of economic planning in India.

9. Describe briefly the various statistical tools used in economic analysis.

10. Write short notes on:

a) Limitation of Statistics.

b) Statistics in Economic Analysis.

S.C.E.R.T. West Library

UNIT II

COLLECTION AND ORGANISATION OF DATA

1. **Statistical Data: Primary and Secondary Data**
2. **Methods of collecting data**
3. **Organisation of Data**

1

Statistical Data: Primary And Secondary Data

Statistics is the scientific methodology of collection, organisation, analysis and interpretation of numerical data. Collection and organisation of data are the first steps in statistical inquiry towards achieving the desired goal. They are the most important steps because all the other steps including one of drawing the conclusion are dependent on them.

Data are obtained through properly organised statistical inquiries conducted by the investigator. Certain terms are used in the process of collection of data and they are explained hereunder:

i) STATISTICAL INQUIRY: By a statistical inquiry, we mean an investigation on a topic by any agency wherin relevant quantitative information is collected. As a result of such inquiries certain numerical facts are gathered about the topic under study. Statistical inquiry also means a search for truth, knowledge or information.

ii) GENERAL PURPOSE INQUIRY: In such a type of inquiry, the data collected can be put to several uses.

iii) SPECIAL PURPOSE INQUIRY: In such a type of inquiry, the data is collected for studying some specific problem.

iv) INVESTIGATOR: One who conducts statistical inquiry is called an investigator.

v) ENUMERATORS: Persons who help the investigator to collect informations are called enumerators.

vi) RESPONDENTS: Respondents are those persons or sources from whom the statistical information is collected.

PLANNING OF A STATISTICAL INQUIRY

When we propose to conduct a statistical inquiry we should keep in mind the following points:-

(A) OBJECT AND SCOPE OF THE INQUIRY: We should be very clear in our minds about the object of the inquiry. This will help us to collect relevant data. In case we are not clear in our objects, we may collect such data which are not at all required for final analysis. Vagueness may also reflect on the design of the questionnaire. A clear idea about the purpose of the inquiry helps in deciding upon the scope or the coverage of the investigation.

(B) SOURCES OF DATA OR INFORMATION: Once the object and scope is determined, the next step would be to determine the sources from which the data have to be collected. The sources of information may be primary or secondary. Information collected by the investigator himself will be called primary data and information obtained from others through published or unpublished sources will be called secondary data. Both the sources i.e. primary and secondary may be used for investigation.

(C) TECHNIQUE OF DATA COLLECTION: There are two methods of data collection: census method and sample method. In the census method, each and every item or variable is studied while in the sample method only a selected number of the items is studied.

(D) UNITS OF DATA COLLECTION: The unit of data collection should be clearly and unambiguously defined. The units should be uniform throughout the study, otherwise they will give misleading conclusions.

(E) DEGREE OF ACCURACY: The degree of acccuracy should be kept in mind. Though cent per cent accuracy is often not possible, yet efforts should be made to achieve a reasonable standard of accuracy. The degree of accuracy depends upon the object of the inquiry.

S.C.E.R.T., W.B. LIBRARY
Date 30.10.98

SOURCES FOR COLLECTION OF DATA

There are different sources from whom data can be collected. The sources can be internal and external. A large number of individual organisations and Government departments generate the data as their regular function which is the internal information. Information collected from outside agency is called external data, which can be obtained from primary source or secondary source. Information from primary or secondary sources can be collected by census or sample method and by conducting surveys and investigations.

COLLECTION OF DATA

SOURCES OF COLLECTION OF DATA

Internal
External

Primary
1. Direct personal interview.
2. Indirect personal interview.
3. Information from correspondents
4. Mailed Questionnaires.
5. Questionnaires filled by enumerators.

Secondary

Published sources
Unpublished sources

Publications of Government, International Bodies
Semi-official publications,
Reports of committees and commissions.
Private publications-
(a) Newspapers
(b) Periodicals
(c) Research papers
(d) Research Institutes
(e) Trade unions
(f) Professional Organisations
(g) Market Reports
(h) Reviews
(i) Reports of companies

INTERNAL SOURCES

A large number of individual organisations and Government departments generate the data for their internal use. They print and publish the information for their own purpose. Generally, this information relates to about the figures of sale, production, salary, wages, profit, loss, dividends etc. Departments compile these data from the available books of the organisation and they send it to their management for further planning and budgeting etc.

EXTERNAL SOURCES

Information collected from outside agencies is called external data which can be obtained from primary sources or secondary sources.

This type of information can be collected through census or sample method by conducting surveys and investigations.

PRIMARY DATA

It is original and first hand information collected by the investigator himself.

SECONDARY DATA

Information collected from the already existing data in the form of published or unpublished reports etc. is called secondary data.

PRIMARY DATA

The data which are collected for the first time by the investigator for his own purpose is known as primary data. For example, to check the blood group of the students of a certain class, the principal asks the school doctors to collect the data. The doctors carry out the blood test on the students and make a record of it in their diary. The data so collected will be primary data for the school.

There are various methods of collecting primary data. The main ones are:-

i) Direct personal investigation;

ii) Indirect oral investigation;

iii) Information through post or mail;

iv) Information through field investigators or enumerators using questionnaire or schedules.

DIRECT PERSONAL INVESTIGATION : This method is also called interview method. The interview may be face to face or on telephone. Under this method, the investigator collects information personally from the concerned sources. He has to be on the spot to conduct the inquiry and has to meet people who are in possession of the required data. It is necessary that in such cases, the investigator has a keen sense of observation and that he is very polite and courteous. He should further acquaint himself with local conditions, customs and traditions so that he is in a positionn to identify himself fully with the persons from whom the information is sought. He should make sure that the persons from whom data is being collected actually know the facts fully and can deliver the goods. He should ask simple and straight forward questions which are capable of being answered precisely and in a language which is not vague but known to the informants.

This method is suitable only for intensive investigations. It involves enormous cost and usually requires a long time. It is naturally not suitable for extensive inquiries where the scope of investigation is wide.

Further in this method the bias or prejudice of the investigator can do a lot of damage as he is solely in charge of the collection of data. This method, however, gives very satisfactory results if the scope of the inquiry is narrow and if the investigator is fully dependable and is completely unbiased.

In this method, the response is very encouraging as most of the persons interviewed are willing to supply information on personal approach.

This method is very commonly used and no doubt gives reliable results in case the investigator possesses certain essential qualities such as efficiency, consistency and tactfulness. In this method, the number of questions asked from the respondent should be limited and strictly confined to the data to be collected, otherwise this method will not give the required results. This method can be safely used:

1. Where too much importance is given to purity.
2. Where the area of inquiry is limited.
3. Where data are to be kept secret.
4. Where importance is given to the originality of the data.
5. Where there is the need of personal interviews by the investigators.

MERITS

1. Accuracy of High Degree- The method generally ensures greater accuracy because of personal touch and supervision of the investigators.

2. Originality - The data are original.

3. Elastic- The method is elastic because the investigator can bring certain changes in the questions according to the need of the situation.

4. Availability of other Informations- He also collects supplementary information about the informant's personal characteristics and investment.

5. Economical - The method is also economical provided the field of investigation is small.

6. Verification of Accuracy - The investigator can also verify the accuracy of the statements given by the informant.

DEMERITS

1. Not suitable for Wide Areas - This method is not suitable where the field of enquiry is vast because the investigator cannot work in a wide area.

2. Personal Prejudice - There are chances of personal prejudice and bias of the investigator as he is the sole incharge of data collection. If he is biased he can draw altogether different conclusions and thus make a misuse of statistics.

3. May not depict real picture of aggregates - Its nature limits its scope. It cannot be used extensively if the inquiry is large and spread over a wide area.

4. More Expensive - It may be very costly where the number of persons to be interviewed is large and they are spread over a wide area.

PRECAUTIONS

While using this method, following precautions should be taken:

1. The investigator should be efficient and polite.

2. The questions should be simple, clear and limited in number.

3. To have accurate data, sincerity should be shown by the investigator.

4. The investigator should keep himself away from personal prejudices.

5. The investigator should be from the same district or state.

INDIRECT ORAL INVESTIGATIONS : In case the direct personal investigation method cannot be used either on account of the reluctance of persons to part with the information when approachedldirectly or on account of the extensive scope of the inquiry or on account of some other reasons, an indirect oral examination can be conducted. In this method, data is not collected directly from the persons concerned but through indirect sources. Under this method, the investigator contacts third parties or witnesses capable of supplying the necessary information. To study the problem about the addiction to alcoholic drinks or drugs, the investigator contacts the shops dealing in these articles rather than the addicts themselves who may not be ready to give correct information. Most of the government committees and commissions make use of this method. Persons who are supposed to possess information on the problems under investigation are invited to give evidence. Such persons are called witnesses and their answers are recorded. It should be ensured that the persons who are being questioned are (i) fully aware of the facts of the problem under investigation (ii) not prejudiced or not motivated to give colour to the facts; and (iii) capable of expressing themselves correctly and can give a true account.

The success of this method depends upon the personal qualities of the interviewers, such as their tact, courage to extract answers and intellectual curiosity and the extent to which they understand the psychological and instinctive reactions of the persons interviewed. Great care and vigilance is needed in assessing the correct value of such information.

Such data collected should not be taken on its face value. Due allowance must be made for the conscious and unconscious bias of the person giving information. The honesty of the interviewer who is not influenced by bribery, nepotism etc. to twist information, is of great importance in the success of this method.

This method is suitable where sources of information are required to be tapped because (i) direct sources do not exist, (ii) cannot be relied upon, and (iii) where the person concerned would be reluctant to part with information.

MERITS

1. Economy - In this method there is the economy of time, money and labour.

2. Convenient - In this method the investigator very conveniently collects the data.

3. Wide Sphere - This method is very useful where the area of inquiry is wide.

4. Quick Work - Work is done very quickly with this method.

5. Less Chance of Favouritism - The investigator cannot provide a biased data.

DEMERITS

1. Lack of Accuracy of High Degree - It is possible that the interviewer may approach those persons who do not possess the relevant information.

2. Partiality of Informants - The witnesses may be prejudicial in favour of or against a particular view point.

3. Lack of Interest of Informants - The correctness of data depend on the informants. The data wll not be correct if the informants are not serious while giving interview.

4. Lack of Uniformity - The data are collected by different persons and from different persons.There is the lack of uniformity in the data.

PRECAUTIONS

The investigator should take certain precautions so as to have reliable information:-

1. Ascertain that the witnesses really possess the knowledge about the subect matter of enquiry.

2. Ascertain that the interviewers understand the psychological reactions of the persons to be interviewed.

3. Ascertain that the witnesses are not prejudicial in favour of or against a particular point of view.

4. Make sure that the interviewers are not subjective in their aproach.

5. Also arrange for a large number of informants.

6. It is very necessary that the investigator should have full faith in the informants.

INFORMATION THROUGH POST AND MAIL :In this method, the investigator appoints local agents or correspondents to collect information from time to time. This method is generally employed by the government departments in those cases where regular information is to be collected from places widely situated. For example, in the construction of whole-sale price index regular information is asked for from different parts of the country. This method is more useful for newspapers, magazines, radio and government agencies when the regular information is desired and where a higher degree of the accuracy is not of much importance. The method enables the coverage on a regular basis of areas which may be distantly situated.

The following points may be kept in view for the success of this method :

1. The correspondents/agents appointed should be trained persons.

2. The correspondents/agents should not be prejudiced.

3. The correspondents should be provided with clear instructions about the concepts and definition to be used in the data collection

MERITS

1. Wide Sphere - When the field of investigation is very vast and information is to be collected from different parts of the country, this method can be used.

2. Economy - Time, money and labour is economised in this method.

3. Relative Accuracy - This method is very useful where relative accuracy is required.

DEMERITS

1. Lack of Originality - The data do not remain primary or original. The correspondents supply the assumed data.

2. Danger of Partiality - The personal bias of the correspondents may enter the reports and to that extent the accuracy is affected.

3. Lack of Accuracy of High Degree - This method cannot be used where high degree of accuracy is required.

4. Time Consuming - For collection of information lot of time is consumed.

5. Lack of Uniformity - Correspondents being large, they give different reports.

PRECAUTIONS

1. Ensure that the correspondents are not prejudiced in favour of or against a particular view point.

2. Make sure that the correspondents are eductated and have a clear picture of the problem.

3. Number of correspondents should be large to get accurate data.

INFORMATION THROUGH MAILED QUESTIONNAIRE OR SCHEDULE : A questionnaire or a schedule is a form containing a number of questions on which the information is sought or required and space is provided for noting down the replies.

The data may be collected either with the help of a questionnaire or a schedule or through personal observations and conversations. Here, we will be mainly concerned with collection of data through questionnaire or schedule.

This method necessitates first preparation of a questionnaire containing a number of questions pertaining to the subject matter under inquiry and suitable space where the information sought can be provided.

Afterwards, a list of persons who are likely to be in possession of the information is drawn up to whom the questionnaire is posted with a request that the same may be returned duly filled in. The method is cheap and results can be obtained therefrom quite expeditiously.

However, in this case, there will always be a large percentage of non-response cases unless it is penalised by legislation. But even legal enforcement is not enough because the persons concerned may not be able to fill in the forms without some assistance which only the enumerators can give. Agricultural statistics in the United Kingdom and the United States of America are collected by this method. In India, this method is employed to collect information regarding annual survey of industries. The chief advantage of this method is that it is less costly as no enumerators are needed. However, the non-response creates serious problem and may even destroy the validity of the results, because it is often associated with the variables which are to be estimated and thereby the results may be biased. The method of mail inquiry would have a chance of success if the persons or organisations from whom the information is collected are not numerous, are enlightened enough to understand the objective of the survey and are willing to cooperate.

MERITS

1. Suitable for Wide Sphere - This method is used where the area is very vast.

2. Economy - This method is very cheap as it involves only the cost of printing the schedules and sending them to the informants.

3. Less Chance of Error - There is less chance of error because the information is given by the informants themselves.

4. Originality - The data are very much original because informants are directly involved in the collection of data.

DEMERITS

1. Lack of Interest of Informants - The greatest limitation of this method is that some informants may not care to fill up the questionnaire. When the number of such persons who have not sent back the questionnaire is very large, it is impossible to draw correct conclusions.

2. Information depending upon questionnaire - If the questionnaire is not simple, wrong answers will be sent by the informants.

3. Effect of Partiality - Personal prejudices may affect the correctness of the data.

4. Fear of Informants - Certain informants may not return the completed questionnaire because of fear.

5. *Effect of complexity of questions* - Sometimes wrong informations are supplied because of complexity of questions.

PRECAUTIONS

1. The questions asked should be simple to understand, short, precise and polite.

2. The object of inquiry must be clearly defined.

3. The interest of the informants be inculcated.

4. There should be a system, a sequence, and proper planning in the questions.

5. The questions should not hurt the informants and be not very personal.

INFORMATION THROUGH QUESTIONNAIRE/SCHEDULE WITH THE HELP OF ENUMERATORS : This method is usually adopted in the case of larger inquiries. In this case, information is collected by the enumerators by interviewing the persons who are likely to be in possession of the required information. This is because even though the questions may be properly drawn up, persons to whom a questionnaire/schedule is addressed may not cooperate either because of "apathy" or for not being able to understand the significance of the questions. On these considerations sometimes enumerators are employed to interview persons who are in possession of information for helping and guiding them in filling up the questionnaire.

In order that the enumerators may be able to properly explain the significnce of the questions to the persons interviewed by them without, in any way, influencing their replies, one way or the other, it is essential that they should be properly trained before being sent out into the field. The enumerators should be explained, the purpose of the survey, the type of information needed and also the definition of basic concepts and rules according to which the size of units must be determined. As for example, in the survey of family income of working classes, the meaning of a family must be clearly explained. Whether brothers residing in the house but with a separate kitchen establishment should be considered as members of the same family; whether friends or other relatives staying with a family and having meals with the family should be treated as a part thereof and so on ?

Do's and Don't for the Enumerators

Do's for the Enumerators:

(a) They must have full knowledge of the scope of the investigation and must be familiar with the terms and codes used for extracting answers.

(b) They must create a friendly atmosphere with the respondent by explaining to him the purpose and necessity of the collection of data.

(c) They must ask questions in the language known to the respondent.

(d) They must record the answer quickly before the next question is asked.

(e) They must explain in simple language the question in case the respondent is unable to follow the question.

(f) They must thank the respondent after recording the full information.

DONT'S for the Enumerators:

(a) They must not interview the respondent in the company of his friends, because privacy is most essential.

(b) They must not frustrate the respondent by insisting on replying the questions when the respondent is unwilling or in a hurry.

(c) They must not suggest the answers.

(d) They must not overwhelm the respondent with their personality.

SUITABILITY OF THE METHOD

This method is most suitable where adequate finance is available, trained enumerators are available to cover a wide field and some significance is attached to the accuracy of the results obtained.

MERITS

1. Suitable for Wide Sphere - This method is used where the sphere of enquiry is wide.

2. Accuracy of High Degree - It can be adopted in all those cases where high degree of accuracy of the collected data is required.

3. Less Time Consuming - It consumes less time.

4. Personal Relations - Since the information is personally collected by the enumerators, there is little chance for non-response or lack of response.

5. Limited Chance of Partiality - There is very limited possibility for partiality or biased information. Responses of illiterate respondents can also be obtained.

DEMERITS

1. Costly Method - This method is quite costly as enumerators are generally paid persons.

2. More Time - This method is very much time consuming.

3. Requires Trained Enumerators - Where trained enumerators are not available, data collected by them may not be reliable.

4. Effect of Partiality - The enumerator might be a biased person and may not truthfully enter the answers given by the respondents.

5. Effect of inability of enumerators - The data collected may not be accurate because of the inefficiency of enumerators.

PRECAUTIONS

1. The enumerator should be able, and efficient.

2. The enumerator must be given a filled up questionnaire in the form of a sample.

3. The questions should be simple and clear.

4. The enumerators should be educated.

5. The investigator may himself check a certain percentage of cases by direct interview.

6. The enumerator should be conversant with the local conditions.

7. The enumerator must be properly trained.

This method is mostly used by the government. For conducting population census all over the world this method is used. In other fields the method is not quite popular because it is too expensive.

DRAFTING A QUESTIONNAIRE

Whatever procedure is used in collecting information through questionnaire or schedule method, its meaning and value depends upon the way in which the specific questionnaire or schedule is prepared. The drafting of a questionnaire which will yield the desired information in unambiguous form is often a difficult exercise. Even investigators with long experience find it worthwhile in initiating a new field of study, to make up a preliminary draft of the questionnaire, actually use it in a limited survey and then on the basis of the results thus gained, make up a revised set of questions for the main inquiry.

The basic requirements of a questionnaire are that the questions should be clear as to the information sought. They should be concerned with facts that the respondents are able and willing to furnish.

The following points should be kept in mind while preparing a good questionnaire:

(i) Questions should be simple, direct and clearly worded. There should be no ambiguity in respect of terms and concepts used.

(ii) The number of questions should be kept to a reasonable limit and consistent with the field of investigaion. Too few questions may not throw light on the problem under investigation and on the other hand too many questions may frustrate the respondent who may start giving haphazard answers to the later part of the questionnaire/schedule.

(iii) The questions should not be such which may in some way offend the respondent. Every care must be taken to extract answers of personal questions. The questions regarding the social status, financial position and educational attainment should be carefully worded because most of the people are very sensitive in replying to such questions.

(iv) The questions should be arranged in some logical order with sufficient space for answers in the questionnaire.

It is essential to enclose a letter with the questionnaire to the respondent in which it should be clearly stated as to what is the purpose of collecting the information, supplied information will be strictly confidential and will not in any case be used against the respondent.

The questions may be divided into four broad types:

(a) Simple alternate questions - In this type, the answer to the question is either (1) Yes / No or (2) Yes / No / Do not know. For example, do you own a house in Delhi?

Is the rent of an accommodation in Delhi cheaper than Bombay? Yes / No/ Do not know.

(b) Multiple choice questions - In this type, the answers to the questions may have many alternatives.

Example :

What is your marital status?

(i) Married (ii) Unmarried (iii) Widowed (iv) Divorced (v) Separated

(c) Open type questions - In this type of questions, the respondent may have different opinions about the problem of the inquiry, e.g.

What is your opinion about the government policies about agricultural production in India?

What is your opinion about the continuation of 10 + 2 + 3 system of education in India?

(d) Specific answer type questions - In such questions, the respondent has no choice to offer but to give the exact specific answer e.g. What is your name?

What is your occupation? and so on.

SPECIMAN OF A GOOD QUESTIONNAIRE IS GIVEN BELOW

NATIONAL COMMISSION ON SELF-EMPLOYED WOMEN

QUESTIONNAIRE

1. Identification

1.1 State ____________ ☐☐☐☐☐

1.2 District ____________ ☐☐

1.3 Town/Village____________ ☐

2. Respondent's Particulars

2.1 Name__________ ☐☐☐☐☐☐☐☐☐

2.2 Religion______________ ☐

1. Hindu, 2. Muslim, 3. Sikh, 4. Jain,
5. Buddhist, 6. Parsi, 7. Christian,
8. Others.

2.3 Caste/Community Specify__________ ☐

1. SC, 2. ST, 3. Backward Classes,
4. others

2.4 Age______ ☐ 2.5 Education________ ☐

2.6 Do you work ? Yes ☐

No ☐

2.7 Are you member of any organisation ?

1. No 2. Member only, 3. Office bearer ☐

If yes, name of the organisation__________

2.8 What work the otherwomen in the neighbourhood/mohalla / village do ?

___

___

___

3. Particulars of Family Members		Total	Literate	Earner
3.1 Adults (18 years and above)	Male			
	Female			
3.2 Children (below 18 years)	Male			
	Female			
3.3 Total Members	Male			
	Female			

3.4 Total annual earnings of family in Rs. ☐

3.5 Total annual earning of respondent in Rs. ☐

4. Nature of Activities engaged in

Agriculture ☐ Horticulture ☐

Rearing Live-Stock ☐

Crafts (Processing and Manufacturing) ☐

Collection of Goods ☐

Services ☐ Vendors/Hawkers ☐

Construction Workers ☐

5. What problems do you face in your work ?___

Date, Name and Signature___

of the interviewer ___

METHODS OF COLLECTING PRIMARY DATA
COMPARATIVE ANALYSIS

Method and situation in which to use the method	Merits	Demerits	Precautions
(1)	(2)	(3)	(4)
1) Direct personal Investigation: has to contact the source of information directly and personally. Following are some of the situations where this type of investigation can be possible: (a) When the area of investigation is limited, (b) When a higher degree of accuracy is needed, (c) When the results of the investigation are to be kept confidential, (d) where the area of investigation is complex and heterogenous, and (e) Where a greater degree of intellectual power, experience and labour are required.	1) Original data is collected.	1) It can be used only when the area of investigation is very small.	1) In the use of this method, precautions must be taken in regard to the nature, behaviour and personality of the investigator.
	2) Reliable information is obtained.	2) It is more expensive and time-consuming.	2) The investigator must possess the aptitude necessary for conducting personal interviews.
	3) There is uniformity in the collection of information.	3) The chances of personal bias are more.	3) The questions should be limited in number.
	(4) Investigator can use his personal influence and choice as the situation demands.	(4) It is more complex and tells heavily upon the resources of the investigator.	(4) Only relevant questions should be asked.
	(5) Provides more accurate results.	(5) X	(5) X

(1)	(2)	(3)	(4)
(2)Indirect Oral Investigation: There are certain situatiuons when the informants cannot be approched directly. In such situations we have to collect data from those persons who may possess some knowledge about the inquiry under investigation.Such persons are known as 'Witnesses'. This method is suitable in the following situations: (a)When the area of investigation is large, (b) When the information cannot be obtained directly from the informants. (c) When expertise in collection of data is required to be used. The government and other such institutions generally make use of this method.	(1) Larger area can be covered (2) It is simple and convenient method of investigation. (3) It is economical and saves time, money and labour. (4)Morecomplex data can also be analysed. (5) It is not open to bias. (6) It is possible to obtain complete and adequate information regarding the inquiry.	(1) The results may not always be true and accurate. These may be open to suspicion and doubt. (2) The informants may not be serious and hence may send carelessly framed replies.	(1) It must be ascertained that the informants possess the required knowledge and are interested in providing the information. (2) The number of informants should be fairly large in order to obtain reliable results. (3) It must be ascertained that informants are not biased and are providing corret and true information only.

(1)	(2)	(3)	(4)
(3) Information through Correspondents: In this method the investigator does not collect the information directly from the respondents. The task is handed over to a few persons or a group of persons who are known as correspondents.This method is more useful for newspapers,magazines, radio and government agencies to collect regular information of higher accuracy.	(1) It is a very economical method. (2) It is very convenient also. (3) Extensive areas can be covered by the investigator.	(1) There is lack of originality in data. (2) The information can be biased. (3) Degree of accuracy is generally very low. (4) Information made available may be only out-of-date.	(1) Only properly trained correspondents should be appointed. (2) Efforts should be made to reduce the personal bias of the correspondents. This bias may arise due to cast,religion, etc.
(4) Information through mailed questionnaire or schedules: A questionnaire is prepared and sent by post to various persons who are likely to be in possession of the information required and they are requested to return the same duly filled in. Generally	(1) Suitable for wide sphere. (2) Method is very cheap as it involves only cost of printing and postage.	(1) Lack of interest of informants. They may not care to reply. (2) If questionnaire is not simple, wrong answers may be supplied.	(1) Questions should be simple, easy to understand , precise and polite. (2) Object should be clearly defined.

(1)	(2)	(3)	(4)
prepaid postage stamps are used to facilitate the informants to post without incurring any cost.	(3) Less chance of error because the data is supplied by the informants.	(3) Personal prejudices may affect the correctness of data.	(3) Proper planning in setting questions.
	(4) Originality.	(4) Fear of informants.	(4) Questions should not hurt the informants.
(5) Information through Questionnaire or Schedule filled in through Enumerators. Enumerators interview the informants on the basis of a questionnaire or a schedule. Answers are filled in by enumerators. It covers wider sphere.	(1) Suitable to cover inquiry of wide sphere.	(1) Very costly.	(1) Train the enumerators.
	(2) Accuracy of high degree.	(2) Requires training of enumerators.	(2) Questions should be simple and clear.
	(3) Consumes less time.	(3) May be affected by bias of enumerators.	(3) Enumerator should be educated and know the job.
	(4) Personal relations.	(4) Inefficient enumerators may make it less effective.	(4) Enumerator may himself check a certain percentage of questionnaire
	(5) Less fear of partiality.	X	X

SECONDARY DATA

According to M.M.Blair "Secondary data are those which are already in existence and which have been collected for some other purpose than the answering of the question at hand".

Sometime it is practically not possible to collect the primary data and thus we have to take the help of secondary data, e.g., to find the birth rate or death rate of different districts in Punjab we write to the Municipal Authorities of cities and towns who have already recorded the birth and the death of the persons in their jurisdiction and on the basis of the information so collected we decide about the birth and death rates. Similarly in order to compare the cost of living index number of Bombay, Calcutta and Delhi we may write to the Government *agencies* in these cities and collect the data from them.

The main sources of the secondary data are;

(a) Published sources and (b) Unpublished sources.

(a) PUBLISHED SOURCES: Various international, national and local agencies regularly publish statistical data. Certain periodicals incorporating statistical data are published by private agencies also. The following are the most important sources for the published data:-

(i) INTERNATIONAL SOURCES: All Foreign Governments besides international agencies publish regular reports on their economy and other related topics of internaional significance. The reports of the following organisations are published regularly for the members:-

1. United Nations Organisation (U.N.O.)
2. International Labour Organisation (I.L.O.)
3. International Atomic Energy Agency (I.A.E.A)
4. Food and Agriculture Organisation (F.A.O.)
5. United Nations Educational, Scientific and Cultural Organisation (UNESCO)
6. World Health Organisation (W.H.O)
7. International Bank for Reconstruction and Development(I.B.R.D.)
8. World Meteorological Organisation (W.M.O.)
9. Inter-Government Maritime Consultative Organisation and many others.

(ii) NATIONAL SOURCES: The Central Government and the State Governments and their different departments and undertakings publish their different reports regularly on their economic activities. The reports are more reliable and authentic. Some of the important publications are :-

1. Indian Trade Journal (Weekly).
2. Statistics of Foreign Trade (Monthly).
3. Indian Customs and Central Excise Tariff (Annual).
4. Reserve Bank of India Bulletin (Monthly).
5. Report on Currency and Finance (Annual).
6. Basic Statistics relating to Indian Economy (Annual).

7. Bulletin of Agricultural Prices (Weekly).
8. Indian Forest Statistics (Yearly).
9. Indian Labour Year Book (Yearly).
10. Census of India (Ten Yearly).
11. Annual Abstract of Statistics and so on.

(iii) SEMI-OFFICIAL PUBLICATIONS: Semi Government organisations like Corporations, municipalities, districts boards, block samitis, taluka samitis and panchayat samitis also publish reports in respect of birth and death, health and sanitation, education and many other related topics.

(iv) COMMITTEES AND COMMISSIONS: The Central Government or State Government appoint several committees or inquiry commissions to find out their expert views on matters of great economic importance or related matters. Reports of such committees/commissions are of great significance and provide invaluable secondary data.

(v) PUBLICATIONS OF THE PRIVATE RESEARCH AND COMMERCIAL INSTITUTIONS: The following commercial (Financial) Institutions and Research Institutions publish their reports regularly:

1. Indian Chambers of Commerce and Industries.
2. Stock Exchanges.
3. National Council of Educational Research, and Training.
4. Indian Statistical Institute.
5. National Council of Applied Economic Research.
6. Institute of Economic Growth, etc.

(vi) NEWSPAPER AND MAGAZINES: Certain Newspapers and magazines collect data in respect of many social and economic aspects. Practically all newspapers/magazines have their research bureau to collect data. To name a few of them are:

(1) Economic Times,

(2) Financial Express and others.

(vii) RESEARCH SCHOLARS: The individual research scholars of the universities collect data for completing their research and these data are published with their research papers.

(b) UNPUBLISHED SOURCES: There are certain records

maintained properly by the government agencies, private offices, firms and others which are not published.

LIMITATIONS OF SECONDARY DATA

One should use the secondary data with care and full precaution and should not accept them at face value as they may be suffering from the following limitations:-

(a) They may not have been collected by proper procedure.

(b) They may not be suitable for the required purpose. The information which was collected before may not be suitable and relevant for the present inquiry.

(c) They may have been influenced by the prejudices of the investigator.

(d) They may be out of date and not suitable in the present circumstances and conditions.

(e) They may not be reasonable and of desired standard of accuracy.

(f) They may not cover the full period of our investigation.

PRECAUTIONS IN THE USE OF SECONDARY DATA

The investigator should consider the following points before using the secondary data:-

1. Whether the data are reliable ?

2. Whether the investigator was sincere about his investigation ?

3. Whether the method adopted to collect the data was suitable and proper ?

4. Whether the data are suitable for the purpose of the present investigation ?

5. Whether census method or sampling method was used while collecting the data ? If sample method was used, how much was the size of the sample ?

6. What degree of accuracy was desired by the agency ? How far was it achieved ?

7. How long was the time of enquiry ? If they have been collected in short period, then they will be more reliable and vice-versa.

8. What was the policy about approximation ? Lesser the degree of approximation, more will be the correctness of the data.

9. Which methods were used by the investigator and upto what

extent they were reliable ?

10. From which source were the data collected ?

11. Upto what extent old data were used ?

12. Whether the data collected before was subject to some tests to find out the reliability of the data ?

13. What was the purpose of the agency that publishes statistics? Whether it is in a routine manner or particularly published for a particular purpose. The purpose may be general or specific, restrictive or inclusive.

14. What was the type of population to whom they were addressed ?

15. In what units the data are expressed ? Are these the same at different times, at different places, and for all cases at the same time or place ?

16. Are the data pertaining to homogenous conditions ? Prof. Bowley has rightly pointed that "*It is never safe to take published statistics at their face value without knowing their means and limitations, and it is always necessary to criticise arguments that can be based on them*". Thus the secondary data should not be used at its face value. It is risky to use such statistics unless they have been properly scrutinised and found reliable, suitable and adequate.

DISTINCTION BETWEEN PRIMARY AND SECONDARY DATA

The distinction between primary and secondary data can be determined on the basis of the criteria as shown in the table below :-

Basis	Primary Data	Secondary Data
1. Originality	It is original, because the investigator himself collects it.	It is not original. Someone collects the data and someone else uses it.
2. Expenses	It requires large expenses in terms of money, time and energy.	It is relatively a less costly method because data is already collected.
3. Suitability	It is collected according to the needs of the researcher.Thus suited the most.	The aim may be different while collecting such data. May not be suitable to the objects of inquiry.

Basis	Primary Data	Secondary Data
4. Prejudice	Possibility of personal prejudice in its collection by the investigator.	Possibility of personal prejudice may be of lesser degree.
5. Accuracy	Information obtained by this source will be more accurate.	Information may not be quite accurate.
6. Precautions	No extra precautions needed in making use of this data.	It should be used with care.

QUESTIONS

.. Distinguish between external and internal sources of data.

2. Distinguish between primary and secondary data. Illustrate your answer with examples.

3. How does a "Mailed questionnaire" differ from a "Schedule" used in the collection of data through interview ? Explain and illustrate.

4. Write short notes on :-
(a) Primary data.
(b) Characteristics of a good questionnaire.

5. What is Secondary Data ? Describe the various sources of collecting secondary data.

6. Describe the questionnaire method of collecting primary data. What precautions must be taken while preparing a questionnaire ?

7. Briefly describe:
(a) Kinds of Statistical inquiry.
(b) Investigtor.
(c) Respondents.
(d) Enumerators.

8. Point out the major problems in constructing schedules and questionnaires.

9. Enumerate the various methods of collecting primary data. Discuss the merits and demerits of any one of them.

10. What do you mean by statistical inquiry or investigation ?

11. What points should be kept in mind while planning for statistical investigation or inquiry ?

12. Discuss the comparative merits of various methods of collecting primary data.

13. What is a questionnaire ? Give a specimen of a questionnaire.

14. What precautions should be taken before using secondary data ?

15. Discuss the limitations of secondary data .

2

Methods Of Collecting Data

Collection of data is the first step in any statistical investigation. There are two methods of collecting data. They are:

a) Census Method or Census Survey; and

b) Sample Method or Sample Survey.

Before discussing these methods, the terms " Population" and "Sample" need to be clearly understood.

The term "Population" when used in statistics has a different meaning. Population here refers to the aggregate of items on which measurements are made in any statistical inquiry. It does neither mean number of people living in an area nor does it mean the entire world. When during the course of investigation, we consider all those persons or units or items , it is called taking the entire population for analysis and inquiry. For example, we want to know the average height and weight of all the boys studying in Central Schools of Delhi in class XIth. Suppose there are 10 Central Schools in Delhi and about 2000 boys are studying in XIth class in these 10 schools. If we weigh and take height of all the 2000 boys and then compute average height and weights from this data , it will be called taking average from the population or universe.

Sample: However, if we want to save labour and time, we can take 20 boys from each school and then take their height and weight and calculate average from 200 boys, and interpret the result, it will be called taking average from the sample.

Thus sample refers to a part of the population selected for analysis, the part is the representative of the whole population. Examples of sample surveys in our daily life are:-

a) A housewife by taking a few grains of rice from the cooker where rice is being boiled, decides whether the rice in the cooker have been fully boiled or not.

b) You generally taste one or two grapes before buying grapes from the fruit seller.

c) A teacher asks a few questions to a student to test his intelligence.

d) A doctor after testing a few drops of blood taken out from the body can tell about the blood group or the ailment etc.

Thus, sample is a part of a larger group and we generally decide about the complete group on the basis of sample drawn from the same group.

METHODS OF COLLECTING DATA

As stated above, there are two methods of collecting data viz. (a) Census Method and (b) Sample Method.

CENSUS METHOD

Under this method, the data is collected from every individual or item of population under investigation. In this method, the results are derived from a study of the complete universe. For example:

i) We have to interview each and every parent whose ward is studying in one or the other schools in Delhi in order to study the problem of schooling expenses of students studying in the schools of Delhi.

ii) The businessman would like to assess the demand for goods produced and services rendered and plan the production targets for the future on the basis of the data collected from the population using the goods manufactured by his concern.

A large number of investigators are to be trained to collect data from the population and a big organisation is required to carry out the investigation. That is why, only the Government uses this method to carry out census of population, production, landholdings etc. In India the census of population is based on this method.

This method is generally used in the following cases:

i) When the field of investigation is limited.

ii) When more accuracy is desired.

iii) When money spent and time taken for collecting the data are of no consideration.

This method is considered to be more reliable because the results are obtained after studying each and every individual item of the population.

MERITS OF CENSUS METHOD

This method has the following merits:

1) INTENSIVE STUDY:-Since information is collected from each and every unit in the population intensive study is possible under this method.

2) *ACCURACY OF HIGH DEGREE:-*As the entire population is involved in the process of investigation, accuracy of the highest order is assured.

3) *SUITABLE IN CASE OF HETEROGENEOUS UNITS :-* This method is most suitable where the units are heterogeneous i.e. different from one another.

4) *ESSENTIAL IN CERTAIN CASES:-* This method is a must in certain cases. Take the case of population census. We cannot draw correct conclusions if the sample method is used.

DEMERITS

This method suffers from the following disadvantages or demerits:-

1. *COSTLY* :This method is very costly as it involves a good deal of money, time and energy. Individuals or small organisations cannot resort to this method.

2. Only rich people or big organisations alone can use this method. This method cannot be used by all.

3. This method cannot be used under all circumstances. This method is not useful where the area of inquiry is very vast and the problem is very complex.

4. This method is more time and labour consuming.

5. As the data involved is large, there are possibilities of statistical errors creeping in.

LIMITATIONS

This method has the following limitations:-

1. This method is practically not possible in certain cases e.g. when we want to purchase one kilogram of grapes, it is not feasible to taste the complete lot to test the quality of grapes.

ii) Due to maximum consumption of time, sometimes, the data collected becomes obsolete and loses its consistency.

iii) Follow up is very difficult because it is very difficult to carry out the investigation again in case of any doubt in the result obtained.

SAMPLE METHOD

Under this method, a small group is selected from a larger group. In other words, we study a small portion of the population and not the complete population e.g. to study about the pocket expenses of the students of a certain Government School, we study about the pocket expenses of a few students and the results obtained are generalised for the students of the whole school. The most important characteristic of this method is that the selected sample should be the representative of the whole population.

Sample method is considered to be the best method in most of the statistical investigations, because of the limitations of the census method. It is also sometimes not possible to have the complete enumeration of the whole population. We have to depend on the sample method in those cases. There are several situations in which we may be compelled to study only a fraction of the population.

We use the sample method in the following cases:

i) WHEN POPULATION IS INFINITE: In case the population is not finite, complete enumeration is not possible. We have to resort to sample method. For example, the act of tossing of a coin will give rise to an infinite population.

ii) FIELD OF INVESTIGATION IS VAST:-This method is also suitable when the field of investigation is very vast or unlimited because this method will consume less time and enumeration can be completed with less amount of resources.

iii) FIELD OF INVESTIGATION LEADS TO FATAL ENDS:- Certain fields of investigation may lead to the destruction of the properties or goods or individuals, a sample of blood is tested and not the whole blood because testing the whole blood may prove fatal.

iv) FIELD OF INVESTIGATION LEADS TO DEVASTATION: Certain fields of investigation may affect the geographical structure of the universe also e.g. to test the mineral resources of a certain region, a small sample of that region is searched and the decision is taken on that basis. In case, we insist on testing the complete region, the whole region has to be drilled to find out the mineral resources of that region.

v) ECONOMY IS CONSIDERED NECESSARY IN INVESTIGATION:- If a large amount of time, energy and money is to be saved, this method is considered appropriate.

ADVANTAGES OF SAMPLING METHOD

Following are some of the advantages of sampling method in general:

1) *LESS TIME CONSUMING:*Time is saved not only in the collection of data, but because of the smaller size, also in analysing the data. It is very useful in getting quick results. It saves time, expenditure and energy because of limited field of investigation.

2) ECONOMY OF COST: Since it is a cheaper method, it is possible to take a sample after a short period of time. More investigations within the same budget.

3) SIMILAR RESULTS: Since it is a scientific method, and trained investigatiors are employed for the collection of data, so results obtained are accurate and sometimes more accurate than those obtained from a census.

4) MINIMUM ERRORS: In this method, it is possible to check the extent of errors, while in census a large number of errors may creep in and it is almost impossible to detect them. The total error is much less in sampling method. A higher degree of accuracy is possible.

5. FOLLOW UP: Because of the smaller size of the sample and the employment of trained investigators, it is possible to cross check the reliability of the data easily.

6. INFORMATION IN DEPTH: The economy of time and money can allow a sampling method to collect more detailed informaion.

7. INDISPENSIBILITY: In certain circumstances, this method alone is possible or prcticable e.g. buying grapes etc. from a fruit seller, examination of blood group etc.

8. CHECKING CENSUS RESULTS: This method is generally used to check the accuracy of data collected by census method.

9. ORGANISATIONALLY MORE CONVENIENT: It is more convenient because it does not require huge force of investigators.

DEMERITS OF SAMPLING

These are the main demerits of the sampling method:-

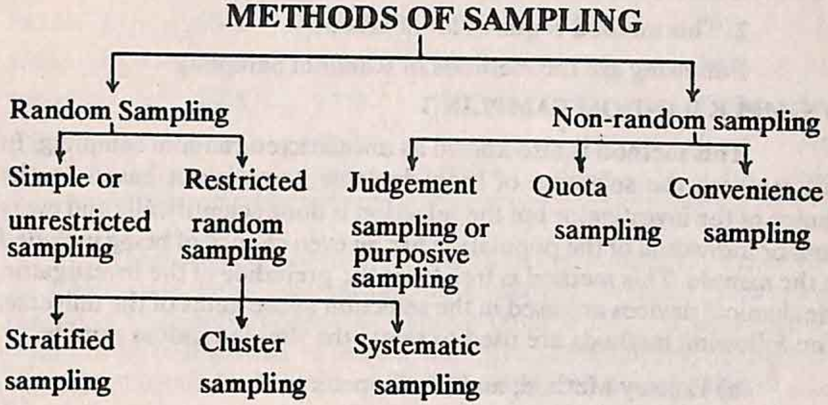
1. The results obtained may be false, inaccurate or misleading if the sample has not been drawn up properly.

2. In the absence of qualified and experienced investigators, the information obtained from the sample cannot be depended upon.

3. On account of personal prejudices of the investigator, the results may not be accurate.

4. This method is not suitable where the units are different from each other and are not homogeneous.

There are a number of methods of sampling, but these can be broadly grouped under two heads, namely, (a) Random sampling and (b) Non-random sampling or deliberate sampling as shown in the chart on next page



Random sampling is of two types, viz. (i) Simple or unrestricted random sampling and (ii) Restricted random sampling. Again restricted random sampling is further classified into (a) Stratified sampling, (b) Systematic sampling and (c) Cluster sampling or multi-stage sampling.

Non-random sampling is classified into:-

i) Judgement sampling or Purposive sampling;

ii) Quota sampling and

iii) Convenience Sampling.

RANDOM SAMPLING

When the samples are selected on the basis of law of probability it is called random sampling. It is also called probability sampling. Each individual in the population has an equal chance of being selected.

MERITS

1) Since the selection of items in the sample depends entirely on chance, there is no possibility of personal bias affecting the results.

2) The selection of items in the sample is objective, so it represents the universe in a better way.

3) No planning for choice is to be drafted.

4) It is less time and money consuming.

5) It provides more accurate data.

DEMERITS

1. Sometimes, it is difficult for the investigator to have up-to-date

lists of all the items of the population to be sampled. To that extent, the use of this method is restricted.

2. This method requires lot of labour.

Following are the methods of Random Sampling:

1) SIMPLE RANDOM SAMPLING

This method is also known as unrestricted random sampling. In this method, the selection of individuals or items is not based on the choice of the investigator but the selection is done scientifically and every item or individual of the population has an even chance of being included in the sample. This method is free from the prejudice of the investigator. Mechanical devices are used in the selection of the items of the universe. The following methods are used to select the simple random sample:

a) Lottery Method; and (b) Tippet's method.

a) LOTTERY METHOD: Lottery Method is the simplest method and is most commonly used even by the ordinary citizen. In this method, chits are made of all the items of investigation and someone who is neutral or impartial is asked to take out the number of chits as are required. The success of this method depends upon the techniques of preparing the chits of same size, shape and colour so that in selection there is no partiality. This method is useful in the following cases:

i) To take out a sample of 15 persons, out of a universe of 150 members, the procedure is to write the name of all the 150 members on separate slips of papers of same size, fold them in the same way, mix them in any container and then ask some neutral person to take out any fifteen slips from the container.

ii) Instead of preparing the chits bearing the number 1 to 150 as done above in case (i), we may have pieces of cardboard, iron sheets, wooden board and write numbers from 0 to 9 as many times as we like and put them in different drums. When these pieces are thoroughly mixed up a neutral person is called upon to take out one piece from each drum. The numbers written on drawn pieces represent the unit's place, ten's place, hundred's place, thousand's place and so on as decided upon. The number thus formed with these pieces is considered a random number. This method is mostly used for State Lotteries, Prize bonds and Premium Prize Bond in India.

*b) Tippet's Method:*The selection of random samples is facilitated by the use of table of random numbers prepared by Tippet, Kendali Smith, Rand Corporation of USA.,Fisher and Yates. Out of these tables, the most important one is L.H.C. Tippet's table which is based on 10400 four digit number so that there are 41600 digits. Given below is the set of first thirty five random numbers:

2952	6641	3992	9792	7969	5911	3170
5624	4167	9524	1545	1396	7203	5356
1300	2693	2370	7483	3804	2762	3563
1089	7691	0560	5246	1112	6107	6008
8126	4233	8776	2754	9143	1405	9025

These numbers have been written out without any order or rule. To select some employees out of 8000 employees, whose names are arranged in alphabetical order we number these employees consecutively from 1 to 8000. If we wish to draw a sample of 25 employees, we make use of Table of Random Numbers. Before starting with the selection of numbers, we decide to go column-wise or row-wise. If we decide row-wise, we select first 25 numbers which are less than 8000 since we have only 8000 employees. From the above table, our selected numbers will be:-

2952	6641	3992	7969	5911	3170	5624	4167
1545	1396	7203	5356	1300	2693	2370	7483
3804	2762	3563	1089	7691	0560	5246	1112
6107							

It is not necessary to start from the beginning, it can be done from anywhere. Now the employees whose serial numbers are mentioned above are selected and this selection will constitute a sample.

2. RESTRICTED RANDOM SAMPLING

It is a method of sampling which is generally used in situations where the population is made of various heterogeneous groups. It is fur*ther sub-divided into:*

a) STRATIFIED SAMPLING:- Stratification means dividing the entire universe into groups. These groups are called strata. The main purpose of stratification is to increase the efficiency of sampling by dividing a non-homogeneous universe in such a way that

(i) there is a homogeneity in the same strata and

(ii) there is difference between the different strata.

After the division of population into groups a required number from each group is selected by a suitable method of simple random samling. For example this method is useful for collecting information concerning the monthly income of the residents of Delhi in the following ways:-

We may first of all categorise the residents according to their income:-

a) having less than Rs. 100.00 per month

b) between Rs. 100.00 and Rs. 500.00 per month

c) between Rs. 500.00 and Rs. 2000.00 per month

d) between Rs. 2000.00 and Rs. 5000.00 per month

e) between Rs. 5000.00 and Rs. 10,000.00 per month

then select a proportionate number from each group and interview them.

MERITS

1) This method is most effective in dealing with highly skewed group such as income data or retail sales.

2) The method ensures better representation of the characteristics of group in the sample.

3) Stratified sampling ensures greater accuracy.

4) Stratified samples are more concentrated geographically thus the time and expenses of interviewing may be considerably reduced.

DEMERITS

1) Subjective judgement used in dividing the universe into strata may affect the accuracy of the results.

2) If the basis of stratification is not properly decided, the results may be misleading.

*b) CLUSTER SAMPLING:-*This sampling is sometimes called area sampling because it is generally applied on a geographical basis. It is also called multi-stage sampling. In this method, the universe is divided into certain number of groups and out of these a few groups are selected at random. Further, selected groups are again divided into sub-groups and from these sub-groups a few are selected and again these sub-groups are further divided into smaller groups and from these smaller groups a few are selected at random. This method can be well explained with the help of the following illustration:-

Let us suppose that we want to select 120 students out of all those being educated by the Directorate of Education, Delhi. We divide all the schools under the Directorate into fifty zones at first stage and take any ten zones out of these fifty zones. Each zone may have eight schools under it and we select any three schools from each of the ten zones selected at first stage. Now each school is divided into four classes, from

9th to 12th and we select any two classes of the schools selected at the second stage. Now each class has forty students and we select any two students from each class selected at the third stage. Thus, we get a sample of 120 students. The selection at every stage is done with the help of simple random selection method.

MERITS

1. This method is used where the area of inquiry is wide.
2. The data collected are accurate.
3. It introduces flexibility into sampling which is lacking in simple method.

DEMERITS

1. This method is very complicated.
2. Less accurate than that of any other method of selecting a sample by a single stage process.

(c) SYSTEMATIC SAMPLING:-Under this method every individual of the population is arranged in alphabetical order or in serial order or the geographical order or in any other order according to convenience. Then the systematic sample is selected by taking every nth item from the universe e.g.1st, 5th, 9th, 13th and so on. In this method also every individual has equal chance of being selected. The first item is selected with the help of simple random method.

NON-RANDOM SAMPLING

If the selection of the individuals from the population to form a sample is not depending on chance but depends on the judgement or convenience of the investigator, then the sampling is said to be non-random sampling.

Under the non-random sampling, the samples are selected with the help of the following methods:

i) JUDGEMENT SAMPLING OR PURPOSIVE OR DELIBERATE SAMPLING:

In this method, the selection or rejection of a certain individual of the population depends upon the investigator. The investigator selects those items which accorrding to him would be representative of the universe,

e.g. the advertisement of the business-man reads: Nine out of every ten interviewed use such and such product, why not you? From this we cannot conclude that any nine persons out of ten selected at random might have used the product. The business-man might have interviewed only those who are his own employees, thus the sample drawn by the businessman is a purposive sample.

The results obtained with the help of this method, are not reliable because the members have been selected with certain purpose either to show the good characteristics of the population e.g. selection of five intelligent students to face the Inspector with the purpose of showing the intelligence of the whole class. This method cannot be in general recommended for any field of investigation.

MERITS

1) This method is widely used in taking business decisions.

2) It is useful in stratified random sampling.

3) This method of taking sample is very easy.

4) When only a small number of sampling units is in the universe, judgement selection would certainly include important elements in the sample.

DEMERITS

1. Under this system personal bias and prejudices have an opportunity of entering the sample.

2. It is not possible to have any idea about the degree of accuracy achieved in any statistical investigation conducted by the method.

2. QUOTA SAMPLING

It is stratified purposive sampling. The actual selection of the members of the sample is entirely left to the investigator who is told how many units having certain characteristics are to be taken, but not which one. The quotas are usually fixed in terms of known strata for instance, of age, sex or occupation. The interviewer is then asked to obtain certain information, say from 60 persons, above the age of 40 and the persons should be 50 males and 10 females. The investigator is free to select the members of the sample provided they come into the specified category.

MERITS

1. This method is mostly used in marketing research studies and public opinion studies.

2. The results obtained are near about satisfactory.

DEMERITS

1. This method is not very popular because it allows some bias and prejudice to enter into the process of selection.

2. The results obtained are not accurate.

3. CONVENIENCE SAMPLING:

This method is quite different from the other two methods of non-random sampling. In this method, a part of the population is selected according to the convenience of the investigator. For example to study the smoking habits of the college students, the investigator may select one or more than one college situated in his neighbourhood.(This method is suitable only when the universe is not properly defined.). The selection is neither based on random nor on judgement but on the convenience of the investigator may be with respect to place due to non-availability of mode of conveyance, or may be with regard to shortage of time at the disposal of the investigator.

MERITS

1) Convenience sampling is often used for making pilot-studies

2) Questions may be tested and preliminary information may be obtained by the chunk before final sampling design is decided upon.

DEMERITS

1) The results obtained by this method are hardly representative of the population.

2) The results are biased and unsatisfactory.

CENSUS AND SAMPLE METHODS

Census Method	Sample Method
1. We take every item of the information for studies.	1. We take a part (Sample) of the information for study.
2. Inferences are drawn from the data of complete universe.	2. Inferences are drawn from the sample regarding the universe.
3. Census method is that in which information regarding every unit of the universe relating to the problem under study is collected.	3. Sample method is that in which information regarding selected number of units of the universe under study is collected.
4. Many difficulties are generally faced when the universe is large.	4. Less difficulties are faced because fewer units of the large universe are observed.
5. A large army of investigators may have to be appointed to collect the data.	5. A small number of trained investigators can collect the complete data.
6. It is difficult to give proper training to a large army of investigators and data collected by them may not always be dependable.	6. A few investigators can properly be trained and data collected by them would be more accurate and reliable.

7. It is expensive and takes more time and needs more energy to collect the data.	7. It requires less money and time. Maximum information about the particular phenomena under study, with minimum money, time and energy is obtained.
8. Only census method can be used in certain cases, such as, for counting the population of a country.	8. Sample method is not possible and reliable in such cases.
9. Census method is impossible in many cases, e.g., in testing the quality of cigarettes, oil, rice, etc.	9. Only sample method is useful in such cases.
10. Universe studies are based on certainties.	10. Sampling studies are based not on certainties, but on probabilities.
11. It requires big organisation to carry the investigation, therefore, only Government uses this method to carry out investigation of population, production, livestock land holdings etc.	11. Sampling can be conducted by small organisation.
12. In case of doubt, it is difficult to undertake the enquiry again for verification. Follow up is difficult.	12 In the case of doubt, enquiries can be undertaken for verification. Follow up is easy.
13. As all the units are covered in the enquiry, highest accuracy can be obtained.	13. If samples are not drawn properly, results may be inaccurate, false and mis-leading.

RELIABILITY OF SAMPLE DATA

The main object of the theory of sampling is to draw conclusion about the population. In other words, the sampling is a tool which facilitates us to generalise the results for the universe on the basis of the results obtained for a sub-group selected from it. It is most essential that the tools testing a certain object must be reliable one.

It is generally inferred that whatever results hold good for the sample, the same must hold good for the population also. The inference is based on the following two important principles:

a) Principle of "Statistical Regularity" ; and

b) Principle of "Inertia of Large Numbers".

These Principles are sometimes called the laws of sampling. In reality these are not the laws of sampling. In a true sense, these are not the laws but characteristics which hold good universally.

a) PRINCIPLE OF STATISTICAL REGULARITY: Accoraing toKing "*The law of statistical Regularity lays down that a moderately large number of items chooşen at random from a large group are almost sure on the average to possess the characteristics of the large groups*". In other words, we can conclude from this principle that if a sample is taken at random from a population, it is likely to possess the same characteristics as that of the population. To have the effective application of the principle, it is most essential that the samples are selected at random. This principle, is, therefore, not applicable to purposive sampling. The principle is ineffective if the size of the sample is too small that the sample may not be truly representative of the population. It is not necessary that the result derived from the sample be in exact agreement with the results to be derived from the population. For example, the average height of the students of Delhi Public School may be 150 cms. by census method whereas it may be 148 cms. or 152 cms. according to sampling method. The sampling results are only exact upto a certain degree of approximation.

b) LAW OR PRINCIPLE OF INERTIA OF LARGE NUMBERS:-By the Inertia of large numbers, we mean the inability of a large mass of data to show an erratic tendency or display "*Sudden Variations which are greater in magnitude*".According toBowley,"*Large quantities are more static than the smaller quantities because the individual item of a collection may increase rapidly whereas the whole collection changes slowly*" e.g. if we throw a dice which is unbiased, the proportion of getting even number would be very nearly equal to 1/2 when the number of throws are sufficiently large, but on the other hand, we may get 8 times an even number or 11 times an even number when a dice is thrown twenty times, thus the result in the proportion of even number will be 8/20 or 11/20 which differs appreciably from that of 1/2. It will be very nearly equal to 1/2 when the number of throws of the dice are unlimited. There may be appreciable deviations, between the expected proportions and observed proportions of the happening of a certain characteristic in case the number of items of individuals are small but these deviations tend to disappear when the sample size becomes sufficiently large. The reasons for this is that the deviations are both positive and negative and their resultant effect does not accumulate but become negligible. As Bowley has said, "*Great numbers have great inertia. It is this consistancy of great numbers that makes statistical measurement possible*".

QUESTIONS

1. Distinguish between a census survey and a sample survey . Give reasons why sample surveys are preferred ?

2. List five important types of sampling methods. Explain the reasons for preferring sample surveys in the collection of Statistics.

(HINT) :- Five important methods of sampling are:-

i) Random Sampling (ii) Stratified Sampling (iii) Systematic Sampling (iv) Purposive Sampling (v) Quota Sampling.

3. . Write Notes on:-

a) Population and Sample

b) Methods of Sampling

c) Stratified Sampling.

4. The Ministry of Education is interested in determining the level of education of unmarried girls in the country. How would you organise a survey for this purpose?

(HINT: First decide whether the survey should be a sample survey or a census survey. Second thing to be decided is about the source of data-primary or secondary. Third thing-if you decide to have sample survey, then the best method would be to prepare a questionnaire, which can be got filled by engaging investigators).

5. Compare the census and sample methods of collecting data with reference to reliability, time involved and costs.

6. Distinguish between random sampling and systematic sampling. Give suitable examples.

7. Discuss the advantages of the sampling method over the census method of data collection.

8. What are the main advantages of sample method ?

9. What is meant by (a) a sample survey and (b) a census ? Why would one often prefer to undertake a sample survey rather than a census ? It is required to conduct a sample survey to estimate per capita expenditure on food in a large city like Delhi. Suggest how a sample of house-hold should be selected for this purpose.

10. Discuss briefly the following :-

a) Random Sampling,

b) Measurement errors,

c) Planning a sample survey.

11. Outline briefly the various steps in planning and conducting a sample survey. What special problems arise in under-developed countries like India?

12. Discuss briefly the following :-

a) Sample Frame

b) Systemtic Sampling

c) Non-Sampling Errors.

13. Name the methods of selecting a sample. Describe the method of stratified sampling with merits and demerits.

14. What do you understand by "Census" and 'Sample" investigation ? Give illustrations.

15. Distinguish between law of statistical regularity and law of inertia of large numbers. Clearly explain them.

16. "Sampling is necessary under certain conditions". Explain with illustration.

17. What are the factors on which the choice between the census and sampling methods depend ?

(HINT: i) The nature and scope of inquiry

ii) Time at the disposal of the investigator

iii) The financial provisions of the inquiry

iv) The degree of accuracy desired and

v) Field of Inquiry).

3

Organisation of Data

According to Prof. Connor organisation or classification of data means "*the process of arranging things (either actually or notionally) in the groups according to their resemblances and affinities, and give expression to the unity of attributes that may subsist amongst a diversity of individuals*". We know that the data collected by the investigator may often be uninteresting. The data may be unorganised and shapeless which can neither be easily compared nor interpreted. Therefore, after collection of the data, the first step is to organise the raw data to present it in some orderly and logical form so that their essential features may become explicit.

The technique of arranging the data in different homogeneous groups is called classification of the data. The process of classification can be compared to the process of sorting out letters, packets, parcels and money orders etc. in a Railway Mail Service (RMS). Just as in sorting out operation, all collected letters, packets, parcels and money orders are separated on the basis of a common characteristic i.e. the destination such as Bombay, Calcutta, Mysore, Patna etc. and they are then put in separate bags. Similarly, in the process of classification, data are classified into various homogeneous groups or classes on the basis of similarities and resemblances. In the school library books of the same subject according to author's names are put together in one shelf. Students in a school are grouped or classified into different sections according to their choice of subjects viz. Medical Sciences, Natural Sciences, Commerce, Arts etc.

Prof. Connor has defined classification as "*An art of grouping the individuals or their characteristics into classes so that all the items of the same related facts are put together*". In other words *it is a technique which divides the items of the universe according to their importance, similarities, dissimilarities or objectives*".

MAIN OBJECTS OF CLASSIFICATION

The following are the main objectives of classification of data:-

1. To reveal clearly the points of similarities and dissimilarities in the statistical data.

2. To arrange the raw data in a scientific manner so that it may be easily understood and interpreted.

3. To eliminate unnecessary details so that the salient features of the collected data become more readily understandable. It minimises mental strain.

4. To be helpful in comparison. If the data are unorganised and dissimilar, it cannot be compared. For proper comparison classification of the data is needed.

5. Classification help in making the data very attractive and effective.

6. To help the investigator and others in forming a mental picture of the phenomenon to which the given data relate.

7. To facilitate the work of tabulation. Classification is the basis of tabulation. Tabulation will be proper, if the classification is proper.

CHIEF CHARACTERISTICS OF A GOOD CLASSIFICATION

An ideal classifiction should possess the following characteristics:

1. The data in a particular group should be homogeneous. Hetrogeneous data should not be taken in that group.

2. The classification should be simple and clear. Ambiguous classifiction would defeat the purpose for which it has been done.

3. The basis of classification should be relevant to the object of investigation. This makes the classification more scientific and more sensible.

4. An ideal classification must possess the feature of comprehensiveness.

5. An ideal classification should be able to adjust itself to new changes and should not be a rigid one. With the changed circumstances some data may have to be dropped while the other may have to be incorporated. The classification should be moulded in such a way that it should be in a position to adopt new changes and at the same time not lose its original worth to suit the nature, object and scope of inquiry.

7. The classification should contain the characteristics of stability. If every now and then the data has to be put into different classes, the stability of the classificaion is lost and comparisons and inferences may not be very satisfactory.

METHODS OF CLASSIFICATION

BASIS OF CLASSIFICATION OF DATA

- Spatial or Geographical classification
- Chronological or temporal classification
- Qualitative classification
- Quantitative classification

1. SPATIAL OR GEOGRAPHICAL CLASSIFICATION

The data is classified on the basis of geographical considerations or locational differences between the vaious items. Under this method data are grouped with reference to countries, states, cities, districts, colonies etc. It is classified either in the alphabetical order for reference or order of size of the value, for immediate comparison.

(a) Alphabetical order (names of the countries)

Country :	America	Canada	China	India	Pakistan
Yield of wheat in kilograms. per hectare in 1986	2312	2240	2997	2032	1881

(b) Descending order (figures in the second row above)

Country.	China	America	Canada	India	Pakistan
Yield of wheat in kilograms per hectare in 1986	2997	2312	2240	2032	1881

In the above example in (b) the figures have been arranged in descending order.

2. CHRONOLOGICAL OR TEMPORAL CLASSIFICATION

When the data are to be classified on the consideration of time, the classification is called chronological classification. The data are grouped according to years, months, weeks etc. Generally it is classified in ascending order of time but in certain cases it may be descending order of time.

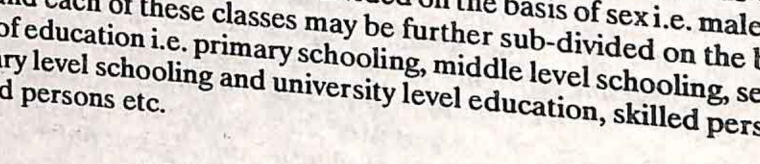
For Example

Year :	1950-51	1960-61	1970-71	1980-81	1985-86	1986-87
Export of tea from India (in crores of rupees)	80	124	145	335	626	550

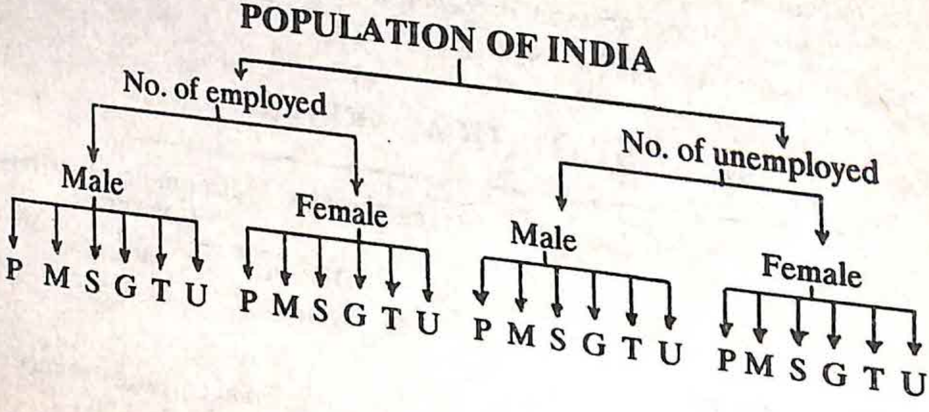
OR	1986-87	1985-86	1980-81	1970-71	1960-61	1950-51
	550	626	335	145	124	80

3. QUALITATIVE CLASSIFICATION

When the data is classified on the basis of some quality or attribute such as (i) honesty (ii) occupation (iii) sex (iv) colour of eye (v) literacy (vi) employment (vii) religion etc. it is called qualitative classification. The qualitative characteristics can not be measured, it can be judged whether a certain attribute is present in the individual or not. For example, to study the problem of unemployment in India we can only find out how many persons are employed and how many are unemployed in the given population of the country. The classification dealing with one attribute is called simple classification.



This classification is also called two fold or dichotomous classificaion. If the classes are further divided on the basis of sex i.e. male and female and each of these classes may be further sub-divided on the basis of level of education i.e. primary schooling, middle level schooling, senior secondary level schooling and university level education, skilled persons, unskilled persons etc.



Here P-Primary education

M-Middle level education

S-Senior Secondary level education

G-Graduation and above

T-Trained/skilled.

U-Untrained/unskilled.

4. QUANTITATIVE CLASSIFIATION

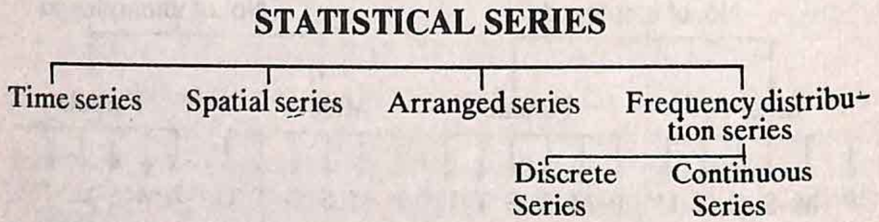
If the collected data is grouped with reference to characteristics which can be measured and numerically described such as height, weight, sales, import, age,income, marks, expenditure, number of persons in the family etc. it is called quantitative classification. An example of the type of classification is given below.

Wages per day in Rs.	20-21	21-22	22-23	23-24	24-25	25-26	26-27
No. of workers	8	10	11	16	20	25	16

Thus, there are 8 workers who are in the wage group of 20-21 per day, 10 workers in wage group of Rs. 21-22 and so on.

STATISTICAL SERIES: When statistical data is arranged in a systematic manner, it is called a *statistical series. Prof. Connor* has defined *statistical series* as, *"If two variable quantities can be arranged side by side so that measurable difference in the one corresponds with measurable differences in the other, the result is said to form a statistical series"*. According to *Prof. H.Secrist, "A series as used statistically, may be defined as thing or attributes of things arranged according to some logical order"*.

TYPES OF STATISTICAL SERIES



1. TIME SERIES

A series of value of some variable arranged according to successive periods of time is called a time series. In such a series the data is presented with reference to time period i.e. year, month, week or day which may be either in ascending or in descending period of time.

EXAMPLE-1

Net national product of India at factor cost at 1970-71 prices.

NET N.P. AT FACTOR COST (AT 1970-71 prices)

Year	Rs. (Crores)
1970-71	34235
1971-72	34713
1972-73	34215
1973-74	36033
1974-75	36590
1975-76	40274
1976-77	40429
1980-81	47414
1981-82	49934
1984-85	57243
1985-86	60143

EXAMPLE-2

Sales of Super Bazar, New Delhi in the last week of December, 1988.

SALES IN SUPER BAZAR NEW DELHI (LAST WEEK OF DEC. 1988)

DAY	SALE(Rs)
Monday	58,192
Tuesday	69,240
Thursday	70,320
Friday	75,560
Saturday	80,890
Sunday	90,520

2. SPATIAL SERIES

A series of value of some variable arranged according to geographical basis or location is called a spatial series.

In such a series, the data is presented with reference to locations, cities, districts, states or countries.

EXAMPLE:

PER CAPITA INCOME 1985

Country	Per capita income in dollars
India	270
Bangladesh	150
Canada	13,680
China	310
France	9,540
Germany Fed. Rep.	10,940
Japan	11,300
Nepal	160
Pakistan	380
Sri Lanka	380
U.K	8,460
U.S.A.	16,690

3. ARRAYED SERIES OR INDIVIDUAL SERIES:

The data as collected by the investigator for statistical enquiry is called *raw data* or unorganised data. This raw data or mass data when put in ascending or descending order of magnitude, is called an *array*.

Suppose an investigator has found the following data on the monthly expenditure on food (in rupees) for 30 house-holds in a locality:-

115	159	196	205	112	223
256	271	310	129	335	169
184	234	245	241	265	298
144	135	172	173	229	243
220	238	278	243	220	238

The above raw data can be arranged in serial order (table No.1) or ascending order (table No.2) or descending order (table No.3).

TABLE NO. 1

ARRANGED IN SERIAL ORDER

S.No	Expenditure	S.No	Expenditure	S.No	Expenditure
1	115	11	196	21	112
2	256	12	310	22	335
3	184	13	245	23	265
4	144	14	172	24	229
5	220	15	278	25	220
6	159	16	205	26	223
7	271	17	129	27	169
8	234	18	241	28	298
9	135	19	173	29	243
10	238	20	243	30	238

TABLE NO. 2

Arranged in Ascending order: *EXPENDITURE ON FOOD IN RUPEES*

112	196	243
115	205	243
129	220	245
135	220	256
144	223	265
159	229	271
169	234	278
172	238	298
173	238	310
184	241	335

TABLE NO. 3

ARRANGED IN DESCENDING ORDER: ***EXPENDITURE ON FOOD IN RUPEES***

335
310
298
278
271
265
256
245
243
243
241
238
238
234
229
223
220
220
205
196
184
173
172
169
159
144
135
129
115
112

4. FREQUENCY DISTRIBUTION

In order to avoid repetition of the variables of the same magnitude we combine them together and arrange them into two columns under the headings (i) Variables and (ii) its frequencies. *Frequency* of the *value of the variable means the number of times of the occurence* of the value of the variable in any one series. For example, the data in table No. 2 above can be shown as :

DISCRETE SERIES

Expenditure in Rs.	No. of house holds (Fr-equency)-	Expend. in Rs.	No. of house holds (Fre-quency)	Expend. in Rs.	No. of house holds (Frequency)
112	1	196	1	243	2
115	1	205	1	245	1
129	1	220	2	251	1
135	1	223	1	265	1
144	1	229	1	271	1
159	1	234	1	278	1
169	1	238	1	298	1
172	1	241	1	310	1
173	1			335	1
184	1				
	10		10		10

FREQUENCY DISTRIBUTION

Before discussing anything about frequency distribution it is very important to know the meaning of certain terms which may be used very often in this type of series.

*(a) FREQUENCY:*The number of times given value in an observation appears is the frequency. For example in the above table Rs. 220/- Rs. 238 and Rs.243/- as expenditures have been incurred by 2 families in each group of expenditures. To be more clear let us take another example. 25 students of your school studying in XI class have the following ages:

15,16,16,17,18,18,17,15,15,16,16,17,15.

16,16,15,16,16,15,17,17,18,19,16,15.

In this case 7 students have 15 as their age, 9 students are of 16 years, 5 students are of 17 years, 3 students are of 18 years and 1 student is of 19 years. We can depict this distribution in a frequency table in the following way:-

FREQUENCY TABLE

Age in years	No. of students (frequencies)
15	7
16	9
17	5
18	3
19	1
	25

*(b) CLASS FREQUENCY:*The number of items belonging to the particular class is called the frequency of that class.

*(c) TOTAL FREQUENCY:*The sum or total of the frequencies is known as the total frequencies.

*(d) FREQUENCY DISTRIBUTION:*The distribution of observations over the several values is called frequency distribution.

(e) CLASS LIMITS:The values of the variables are divided into number of classes. The class limits are the lowest and the highest values of the variable that can be included in that class. The lowest value of the variable in the class is called the lower limit of the class and the highest value is called the upper limit of the class. For example, the class 10-20

means that this class cannot include any item whose value is less than 10 or greater than 20,10 is called the lower limit and 20 is called the upper limit.

(f) CLASS INTERVAL: Class interval is the width of the class i.e. the difference between the upper limit and the lower limit of the class. For example, in the class 10-20 the class interval is 20-10 = 10. The size of the class interval is determined with the help of the number of classes and the range of the data e.g. if the variables varies from 5 to 85 and it is desired to have 8 classes then the size of the class interval is (85-5)/8 = 10

*(g) MIDPOINTS OF THE CLASS:*The point which is midway between the lower and the upper limit of the class is called the midpoint of the class and it is calculated by the formula:

Midpoint of the class = (Lower limit + Upper limit)/2.

For all practical purposes the value of the midpoint is taken as the representative of the class. There are two methods of classifying the data into different classes, (i) exclusive method, and (ii) Inclusive method.

EXCLUSIVE METHOD

In this method the upper limit of each class is the lower limit of the next class. This method ensures the continuity of data. In this method the class intervals can be like this:

0-10, 10-20, 20-30, 30-40 and so on. For example according to this method, the class interval 10-20 will include those values of the variable which are either (i) more than or equal to 10 but less than 20. The value of 20 will be included in the next bigger class i.e. 20-30. Thus, in the exclusive method of classification we include the lower limit and not the upper limit of the respective class.

INCLUSIVE METHOD

In this method the upper limit of a class interval is not equal to the lower limit of the next class. For example, in this method the class intervals are 0-9, 10-19, 20-29 and so on. In this method the continuity is not maintained. In this method both the lower and the upper limits are included in the respective classes.

PRINCIPLES OF GROUPING THE DATA

There is no hard and fast rule for grouping the data but the following general principles may be kept in mind for statisfactory and meaningful classification of data:-

(a) It is advisable to have total number of classes between 5 and 15. The preference for the total number of classes depends on the number and figures to be grouped, the magnitude of the figures,etc.

(b) Odd figures should be avoided for class intervals. The choice should be either 5 or a multiple of 5.

(c) Lower limit of the class should be 0 or a multiple of 5.

(d) For maintaining continuity and correct classes exclusive method be adopted.

(e) The class interval should be equal for all classes.

(f) As far as possible open end classes be avoided.

(g) For frequency distribution we prepare a table having three columns-first for variables second for "Tally Bars" and the third for the total representing corresponding frequency to each class.

PRACTICAL EXERCISES TO PREPARE FREQUENCY DISTRIBUTION FROM THE GIVEN RAW DATA

ILLUSTRATION NO. 1: Prepare a frequency table of the ages of 25 students of XI class in your school:

15,16,16,17,18,18,17,15,15,16,16,17,15,16,16,15,

16,16,15,17,17,18,19,16,15.

SOLUTION: The data is classified below with the help of tally bars:

Age	Tally bars	Frequency
15	~~IIII~~ II	7
16	~~IIII~~ IIII	9
17	~~IIII~~	5
18	III	3
19	I	1
		25

ILLUSTRATION NO.2:A survey showed that the average daily expenditure (in rupees) of 30 households in a city was:-

11,12,14,16,16,17,18,18,20,20,20,21,21,

22,22,23,23,24,25,25,26,27,28,28,31,32,33,36,38.

Arrange the above data in a continuous frequency distribution.

SOLUTION: Here the minimum expenditure is Rs. 11 and Rs. 38 as maximum. We can have the six classes as 10-15, 15-20,20-25,25-

30,30-35,35-40. The data is classified below with the help of tally bars:

CLASSES	TALLY BARS	FREQUENCY
10-15	III	3
15-20	~~IIII~~	5
20-25	~~IIII~~ ~~IIII~~	10
25-30	~~IIII~~ I	6
30-35	IIII	4
35-40	II	2
		30

ILLUSTRATON NO. 3 We are given the following marks secured by 25 students in an examination:
23,28,30,32,35,36,36,40,41,43,44,45,45,48,49,52,53,54,56,56,58,61, 62,65,68.

Arrange these data in the form of a frequency distribution taking class intervals-
20-29, 30-39, 40-49, 50-59, 60-69

SOLUTION:

Class intervals	Tally bar	Frequency
20-29	II	2
30-39	~~IIII~~	5
40-49	~~IIII~~ III	8
50-59	~~IIII~~ I	6
60-69	IIII	4
		25

ILLUSTRATION NO.4 From the following information prepare a frequency distribution using exclusive class intervals :-

MIDPOINT	FREQUENCY
115	6
125	25
135	48
145	72
155	116
165	60
175	38
185	22
195	3

SOLUTION:

The data can be regrouped into a frequency distribution using exclusive method as under:-

Midpoint	Class interval	Frequency
115	110-120	6
125	120-130	25
135	130-140	48
145	140-150	72
155	150-160	116
165	160-170	60
175	170-180	38
185	180-190	22
195	190-200	3

ILLUSTRATION NO. 5. In a hypothetical sample of 20 individuals, the amounts of money with them were found to be:-

Individual	Money (Rs)	Individual	Money (Rs)
1	114	11	131
2	108	12	136
3	100	13	143
4	98	14	156
5	101	15	169
6	109	16	182
7	117	17	195
8	119	18	207
9	121	19	219
10	126	20	235

SOLUTION:

It has not been specially given as to which frequency distribution-discrete or continuous, has to be made. The given frequency distribution can be constructed into a continuous series with the starting value of 80 and with a class interval of 40.

C.I.	Tally Bar	Frequency
80-120	~~IIII~~ III	8
120-160	~~IIII~~ I	6
160-200	III	3
200-240	III	3
		20

ILLUSTRATION NO. 6

Classify the following frequency distribution into another frequency distribution showing the mid-values of the classes:-

INCOME	FREQUENCY
0-100	6
100-200	10
200-300	25
300-400	45
400-500	78
500-600	100
600-700	80
700-800	46

SOLUTION:

C.I.	MIDVALUES	FREQUENCY
0-100	50	6
100-200	150	10
200-300	250	25
300-400	350	45
400-500	450	78
500-600	550	100
600-700	650	80
700-800	750	46

ILLUSTRATION NO. 7

Express the following data in a cumulative frequency distribution of "less than" and "more than" type.

CLASSES	FREQUENCIES	CUMULATIVE FREQUENCIES
20-30	13	13
30-40	11	24
40-50	18	42
50-60	10	52
60-70	6	58
70-80	5	63
80-90	14	77
90-100	12	89
100-110	11	100

SOLUTION

Less than C.F. table

Wages below Rs	No. of Workers
20	0
30	13
40	24
50	42
60	52
70	58
80	63
90	77
100	89
110	100

More than C.F. table

Wages above Rs	No. of Workers
20	100
30	87
40	76
50	58
60	48
70	42
80	37
90	23
100	11
110	0

ILLUSTRATION NO.8

Prepare cumulative frequency of "more than" and "less than" eries from the following data:-

Marks :	0-5,	5-10,	10-15,	15-20,	20-25,	25-30
No. of Students:	7	10	20	13	12	10
Marks:		30-35	35-40			
No. of students:		14	9			

SOLUTION:

MORE THAN SERIES

Marks	Numbers of students
More than 0	95
5	88
10	78
15	58
20	45
25	33
30	23
35	9
40	0

LESS THAN SERIES

MARKS	NO. OF STUDENTS
Less than 5	7
10	17
15	37
20	50
25	62
30	72
35	86
40	95

ILLUSTRATION NO. 9

Convert the following inclusive frequency distribution into exclusive frequency distribution:

MARKS	FREQUENCIES
1-10	4
11-20	6
21-30	18
31-40	16
41-50	6

SOLUTION:

Steps: 1. Find the difference between the upper limit and the lower limit of the next class..

2. Divide the difference found out in (1) by 2. Here the difference between the upper limit of every class and lower limit of the succeeding class is 1. We divide this by 2 and decrease the lower limit of a class by.5 and increase the upper limit of the same class by .5, thus the new group frequency distribution will be:-

Marks	No. of students (frequencies)
.5.-10.5	4
10.5-20.5	6
20.5-30.5	18
30.5-40.5	16
40.5-50.5	6

QUESTION :

1. Define classification and explain its objects and various methods.

2. What are the main objects of classification ?

3. Explain briefly the chief characteristics of a good classification.

4. Write short note on the following:
(a) Spatial or geographical classification.

(b) Chronological or temporal classification.

(c) Qualitative classification.

(d)Quantitative classification.

5. What is statistical series ? What are its various types ?

6. Explain discrete and continuous variables with examples.

7. Explain briefly the following terms:

(a) Individual series

(b) Exclusive and inclusive class intervals.

(c) Simple and cumulative frequency.

8. Define frequency distribution. State the principles required to be observed in its formation.

9.What is an Array ? How is it useful in frequency distribution ?

10. What do you mean by frequency table ?

11. The following are the rents of 60 houses and you are asked to organise the same by grouping these in intervals of Rs. 5/-

41, 43, 41, 51, 60, 50, 48, 47,46, 56, 55, 58, 61, 64, 62, 63, 70, 64, 66, 58, 61, 45, 45, 46 48, 29, 60,70, 44, 45, 43, 51, 45, 55, 57, 54, 70, 72, 49, 67, 62, 48, 65, 58, 49, 73, 71, 55, 59, 74, 44, 66, 62, 40, 56, 67, 73.

12. Prepare a frequency distribution taking a class interval of 2 from the following data:

1, 3, 2, 4, 5, 6, 7, 6, 5, 7, 6, 3, 2, 1, 9, 7, 6, 3, 2, 1, 5, 4, 4, 5, 4, 5, 4, 4, 8, 5.

13. Prepare a discrete series from the following data :

62, 50, 57, 58, 51, 53, 62, 64, 60, 61, 60, 51, 65, 55, 55, 52, 60, 65, 58, 60, 59, 52, 63, 56, 56, 58, 64, 63, 62, 60, 58, 54, 62, 54, 54, 60, 65, 60, 62, 59, 65, 56, 63, 52, 53, 62, 53, 61, 61, 59.

14. Change the following into continuous series and convert the series into "less than" and "more than" cumulative series.

Marks

Mid values	No. of students
5	8
15	12
25	15
35	9
45	4
55	2

15. Convert the following cumulative frequency into frequency distribution.

MARKS LESS THAN	FREQUENCY
10	4
20	10
30	16
40	23
50	40
60	45

16. Convert the following cumulative frequency distribution into frequency distribution.

Marks More than	Frequency
100	0
90	7
80	12
70	20
60	35
50	40
40	49
30	56
20	72
10	75
0	80

17. Convert the following frequency distribution into cumulative frequency distribution :

CLASSES	FREQUENCIES
0-10	3
10-20	6
20-30	4
30-40	16
40-50	11

18. Classify the following frequency distribution into another frequency distribution showing the mid-values of the classes:-

INCOME	FREQUENCY
0-100	6
100-200	10
200-300	25
300-400	45
400-500	78
500-600	100
600-700	80
700-800	46

19. A survey showed that the average daily expenditure of 30 households in a city was:-

11,12,14,16,16,17,18,18,20,20,21,22,22,23,23,24,25,25,26,27,28,28, 31,32,33,36,38

(a) Form a frequency distribution using class intervals:

10-14, 15-19, 20-24, 25-29, 30-34, and 35-39

(b) What per-cent of the households spend more than Rs. 29/- each day ?

20. We are given the following marks secured by 25 students in an examination:

23, 28, 30, 32, 35, 36, 36,, 40, 41, 43, 44, 45, 45, 48, 49, 52, 53, 54, 56, 56, 58, 61, 62, 65, 68.

(a) Arrange these data as a frequency distribution taking class intervals:-

20-29, 30-39, 40-49, 50-59, 60-69.

(b) Prepare cumulative frequency distribution.

21. We have the following data on the monthly expenditure on food (in rupees) for 30 households in a locality:

115, 159, 196, 205, 112, 223, 256, 271, 310, 129, 335, 169, 184, 234, 245, 241, 265, 298, 144, 135, 172, 173, 229, 243, 220, 238, 278, 243, 220, 238.

(a) Obtain a frequency distribution using class intervals:

100-150, 150-200, 200-250, 250-300, and 300-350

(b) What per-cent of the households spend less than Rs. 250/- per month and what per-cent spend more than Rs. 200/- per month ?

UNIT III

Presentation of Data

1. Tables and Diagrams
2. Time series and Graphs

1

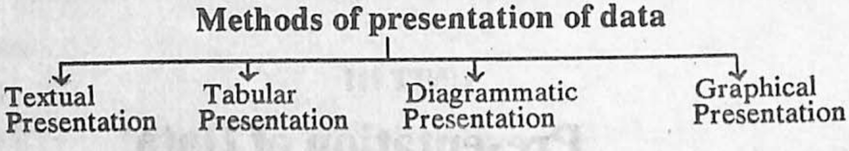
PRESENTATION OF DATA

-TABLES AND DIAGRAMS

Croxton and Cowden say, "*Either for one's use or for the use of others, the data must be presented in some suitable form. Usually, the figures are arranged in table or represented by graphic devices.*" An investigator has to collect data for some study–social, economic, scientific or any other. The collected data is then organised in a systematic way. After the collection of data, it is required to present the quantitative information to expert and non-expert bodies for different purposes. In order to have full worth of time, energy and money spent in the collection of data, it is imperative that the presentation should be done in an easily understandable way. Reading a story or a drama does not have so much effect as seeing a T.V. presentation or seeing a movie of the same story or drama. There are various methods of presentation of data and each method has its own advantages and disadvantages. Let us discuss the various methods of presentation of data.

METHODS OF PRESENTATION OF DATA

There are four methods of presenting the data.



1. Textual Presentation: When a text is presented in the form of paragraphs, it is called textual presentation. The presentation of the statistical data with the help of text is not always effective because the information has to be read again and again to grasp the entire set of figures. It is difficult for a layman to understand the data presented in this way and to single out the desired figures. The main advantages of this representation is that specific emphasis can be laid on certain items which require to be compared. The reader can understand the comparative statements with their explanations.

2. *Tabular Presentation*: Prof. Neiswanger has said, "*a statistical table is a systematic organisation of data in columns and rows.*" The process of arranging statistical data with the help of tables is called tabulation. It is the scientific process of presentation of classified data in a proper order to facilitate comparison. Statistical table is a presentation of quantitative data in vertical columns and horizontal rows.

Usually, the volume of data collected in any statistical inquiry is not always so small that it can be quickly assimilated. Perhaps, one may at best get some very crude idea about the final results. Even so, one cannot be certain about the correctness of these results. Representing the data in a tabular form is really an attempt at condensing the data in a comprehensible form. It also facilitates comparison of data.

Advantages of Tabulation.

The tabulated data can perform the folowing functions:

1. The tabulated data can be more easily understood.
2. The data becomes attractive and leaves a lasting impression.
3. Quick comparison is possible.
4. Tabular arrangement makes summation of items possible.
5. Detection of errors and ommissions becomes easier.
6. The repetition of explanatory phrases and headings are avoided.

7. Space is saved because data are placed in rows and columns.

8. Complex data is simplified.

9. Computation of different statistical measures viz. averages, dispersion and correlation is facilitated.

10. Tabulated data are good for references and they make it easy to present the information on graphs and diagrams.

11. The data is presented in an organised and scientific manner.

12. Tabular presentation simplifies the problems.

Essentials of a Good Table

A good table should possess the following characteristics;

1. The table should be according to the object of statistical investigation.

2. The table should be compact and brief. It should explain only the relevant and desired information needed for further statistical investigation.

3. It should not be overloaded with details.

4. The table should be neat and attractive so that it should appeal to the reader or investigator.

5. The purpose of the table should be clear.

6. Each aspect of the data should be clearly brought out.

7. The table should facilitate comparisons.

8. A table should always make a specific mention of the source of the data, as well as a note about various data gaps.

9. The unit of measuement should be clearly indicated, such as "price in rupees" or "weight in kilograms".

10. There should be no ambiguity in classification.

11. The table should have an appropriate title, captions and stubs, totals, footnotes, etc.

12. The table should show relationship between related items.

13. The table should suit the size of the paper. It should not be a square peg in a round whole.

14. Columns and rows should be numbered when it is desired to facilitate reference to specific parts of a table.

15. Certain figures which are to be emphasized should be in dis-

tinctive type, or in a "circle", or a "box" or between thick lines.

16. A table should have miscellaneous column for the data which cannot be grouped in the classification made.

Parts of the Table

A good table should have the following parts:

(i) The table number: When in a survey or report or book, more than one table is used, it should be numbered in serial order. This helps the reader/user to refer to the particular table for his investigation. The number of the table should be placed in the middle on the top of the table.

*(ii) The title or the heading of the table:*The heading or title of the table should be clearly mentioned at the top. The heading should be brief, clear and concise. It should be unambiguous, well written and self-explanatory. It should be properly punctuated. The table does not have any meaning in the absence of its title. The lettering of the title should be the most prominent of all lettering used in the table. A complete title explains (a) nature of data, (b) time period of data, (c) the field to which the data are related, and (d) basis of classification of data.

*(iii) Captions and Stubs:*Heading and sub-headings given to column are called captions while rows are identified by stubs. Both stubs and captions should be as brief as possible. Columns can be numbered according to requirement. The captions explain the figures in columns and stubs clarify the figures in the rows. We read figures of captions vertically and stubs horizontally.

*(iv) Prefatory note or head note:*It is a statement to explain the contents of the table or main part of it, for example, "All data in million tonnes" etc. It is always placed below the title.

*(v) Body of the table:*It consists of the data or figures to be represented by the table. This is the most important part of the table. It should be made as comprehensive as possible keeping in view the purpose of presentation. There should be proper arrangement of items in columns and rows and irrelevant matter should be excluded.

(vi) Totals: Both row and column totals should be given. Grand totals of rows and columns should be identical and this helps in checking the accuracy of tabulation.

(vii) Footnotes: Footnotes are used to explain the things that may not be clear in the main body. It is a phrase of statements which clarifies some specific items or some specific part of the table or explains the omissions there of and is placed at the bottom of the table.

(viii) Source Note: The source note appears below the footnotes

if they are used. The source notes are indicated in case secondary data is used. Full explanation of the source must be mentioned so that in case of any doubt in the mind of the user, he can refer to the source from where the data has been taken. The source note should give the name of the author, table number, volume number, page number and the place of the publisher of the source.

Specimen of a Table

Table No. XIX.2

PLAN OUTLAY AND INVESTMENT

(Rs. in crores)

Plan	Public Sector Outlay	Investment			Investment % Distribution	
		Public Sector	Private Sector	Total	Public	Private
First Plan (1951-56)	1960	1560	1800	3360	46.4	53.6
Second Plan (1956-61)	4672	3731	3100	6831	54.6	45.4
Third Plan (1961-66)	8577	6300	4100	10400	60.6	39.4
Annual Plan (1966-69)	6625	n.a.	n.a.	n.a.	-	-
Fourth Plan (1969-74)	15779	13655	8980	22635	60.3	39.7
Fifth Plan (1974-79)	39426	36703	27048	63751	57.6	42.4
Sixth Plan (1980-85)	97500	84000	74710	158710	52.9	47.1
Seventh Plan (1985-90)	18000	154218	168148	322366	47.8	52.2

Source: Ministry of Finance, Indian Economic Statistics: Public Finance, December 1986.

Note: Figure for the 6th plan and 7th plan relate to plan-provisions.

Main types of tables

On the basis of construction, tables are of two kinds e.g. (i) Simple table and (ii) Manifold table. Manifold tables can be (a) Double table, (b) Treble table, (c) Manifold table

Simple table

Simple tabulation provides information about one or more groups of independent questions. It is very easy to construct a simple table. This type of table has only two parts.

Table No......

Table showing number of students in various faculties of Delhi University, Delhi-7 in 1988

Faculties	No. of students
1. Arts	
2. Commerce	
3. Science	
4. Agriculture	
5. Law	
	Total:

Manifold Table

It indicates statistical data into more than one category and reveals information about a set of inter-related inquiries. It can be of many types, viz.

(a) Double Table: It provides information about two interrelated characteristics of a particular phenomenon. The table is called two way because it reveals information about two properties. The number of students of the various faculties can be grouped sex-wise-males and females.

TableNo......

Sex-wise number of students in various faculties in Delhi University in 1988

Faculties	No. of students		Total
	Males	Females	
1. Arts			
2. Commerce			
3. Science			
4. Agriculture			
5. Law			
Total:			

(b) Treble Table: In case, we make a study of three interrelated variables, it would form a treble table. The above table can be split up further to provide an additional information e.g., the number of day scholars and the number of hostellers. The table would take the following shape:-

Faculty	No. of students					
	Males			Females		
	Day scholars	Hostellers	Total	Day scholars	Hostellers	Total
1. Arts						
2. Commerce						
3. Science						
4. Agriculture						
5. Law						
Total:						

(c) Manifold Table: These tables reveal information about more than three inter-related questions. The above table can further be split up adding information about married and unmarried at each stage.

Faculty	Males						Females					
	Day scholars			Hostellers			Day scholars			Hostellers		
	M	UM	Total	M	UM	Total	M	UM	Total	M	UM	Total
1. Arts												
2. Commerce												
3. Science												
4. Agriculture												
5. Law												
Total:												

Difference between classification and tabulation.

Both classification and tabulation are the most important activities in the statistical investigation. These help in presenting the data in a proper manner. But, there are differences between the two.

1. The sequence is different in both the activities. First of all, data are classified and then classified data are tabulated. Thus, classification is the basis of tabulation.

2. In classification data are classified on the basis of similarity and dissimilarity. Under tabulation, data are classified in columns and rows. Thus, tabulation is the mechanical function of classification.

Mechanical Tabulation

Now-a-days, tabulation work is done with the help of (i) one or two card punch machines, (ii) sorting machines, and (iii) tabulating machines. This facilitates and speeds up the work. This method is generally used when the study is extensive enough. It saves time. Large survey can be handled conveniently. Accuracy is ensured. Human effort is reduced. Cost is also minimised.

3. Diagrammatic Presentation:

An important object of statistical methods is to reduce the voluminous nature of statistical data and to make these more meaningful and understandable. Classification and tabulation constitute the first step towards the attainment of this objective. But still, such numerically represented material may be quite uninteresting to many of us. Therefore, it is necessary to devise such methods which may present huge mass of quantitative data in a way that is at once comparable and appealing both to the eye and the intellect. For this purpose graphs and diagrams etc. are used for presenting data. The digrams consist of geometrical figures like bars, squares, rectangles, circles or pictures or maps.

Usefulness of the diagrams

Prof. M.J.Moroney has said the following about the usefulness of diagrams-"*Just as maps give us a bird's eyeview of wide stretch of country, so diagrams help us to visualize the whole meaning of numerical complex at single glance*". The usefulness of the diagrams will be clear from the following points:

1. They are *interesting, attractive* and *impressive.* Even a layman can easily get message from diagrams. Diagrams help in publicity and propaganda.

1. They are *simplest methods of presenting data.* They save time and energy of the statistician and the layman. One can understand the figures what they want to say without any strain on mind and knowledge of mathematics as they make the data simple and intelligible.

3. They make *comparison* easy from one period to another period or from one section to another section. Diagrams help in quick comparisons.

4. They have *universal utility.* Since the figures are made very attractive by the presentation of diagrams, they are popularly used in exhibitions, fairs, journals, newspapers, board meetings etc. They are used to supply information to common man. They are widely used in economic, social, business and other fields.

5. They create *lasting impression* in the mind. Thus, they live long.

6. They save *time and labour* because they are easy to understand. They save time which is otherwise needed to draw results.

Limitations of diagrammatic presentation

While using diagrams, the following limitations should be taken care of:

(a) It is very difficult to show minor differences with their help. Construction of diagrams requires relatively more time and expertise.

(b) In case of only one diagram, comparison is not possible.

(c) The diagrams show only a limited amount of information while tabulation can show much information about the data.

(d) By looking at the diagrams, we can not draw correct conclusions. They show only approximate values. The real statistical value is suppressed.

(e) The diagrams may be used for grinding one's own axe during propaganda and advertisement. It is subjective in character. Its interpretation varies from person to person.

General rules for constructing diagrams

The following general rules may be kept in mind while constructing diagrams:

1. A diagram should be *appealing* to the eye.

2. It should be *attractive, neat and accurate.*

3. *A suitable heading* or title on the top of a diagram should be provided.

4. *A proper scale* should be selected to suit the size of the paper.

5. A diagram should be drawn with the help of *geometric instruments*.

6. *Footnotes* may be given just below the diagram to clarify certain points.

7. An *index* must be given for identification of different shades or colours used to make the diagram more attractive.

8. It should be *easy and simple* to understand.

9. The *size* of the diagram should be neither too big nor too small. It should be very much attractive.

10. The important points and data should be shown with *dark colour.*

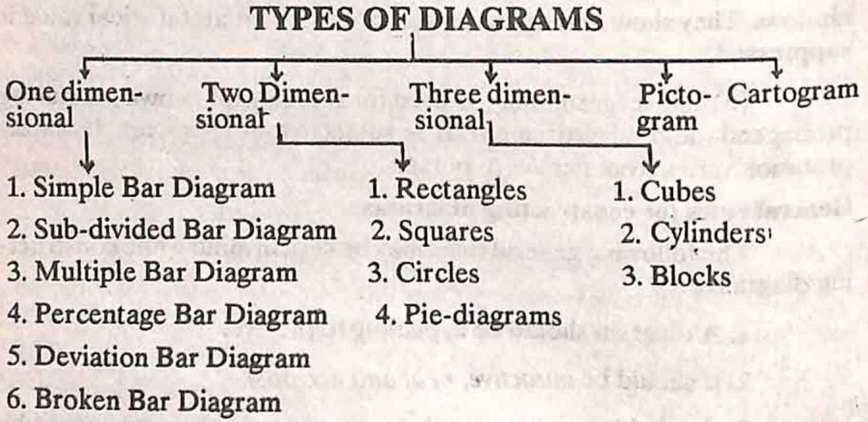
11. *Accuracy* of the data presented should be maintained.

12. The *nature* of the data, *time* of observance of the data, place of occurence of the data and the source of the data must be clearly stated.

13. The diagrams should *not contain too many details* to avoid any confusion.

TYPES OF DIAGRAMS

The following are some of the commonly used type of diagrams:



TYPES OF DIAGRAMS

One dimensional
1. Simple Bar Diagram
2. Sub-divided Bar Diagram
3. Multiple Bar Diagram
4. Percentage Bar Diagram
5. Deviation Bar Diagram
6. Broken Bar Diagram

Two Dimensional
1. Rectangles
2. Squares
3. Circles
4. Pie-diagrams

Three dimensional
1. Cubes
2. Cylinders
3. Blocks

Pictogram

Cartogram

Let us illustrate below these various diagrams.

One Dimensional Diagrams:

(A) Simple Bar Diagrams.

They are very simple to present and only one type of variable can be presented. A simple bar diagram can be drawn either on horizontal or vertical base, but bars on horizontal base, (showing vertical bars) are more common. This type of diagram can present information of one category either in years, months, weeks etc. or relating to different places, sections or groups. All the bars can be beautified by a single colour or shading to make them more attractive.

Illustration No. 1

Represent the data given in the following table with the help of a bar diagram.

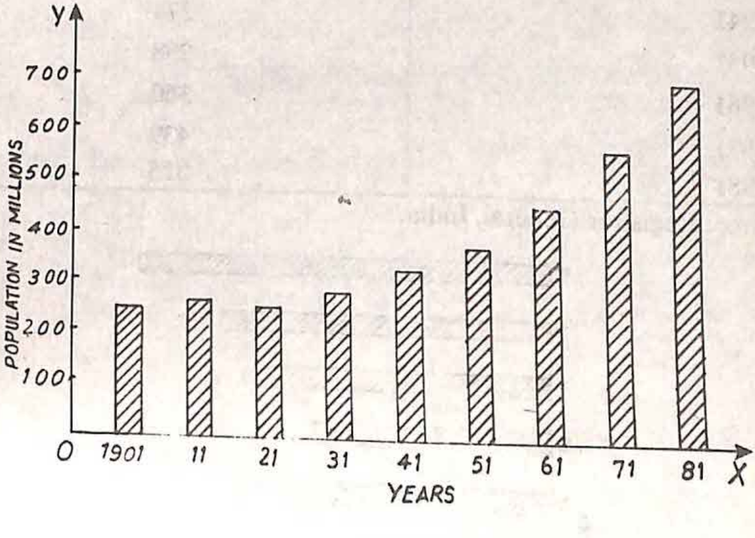
Table

Year	Population in millions
1901	238
1911	252
1921	251
1931	279
1941	319
1951	361
1961	439
1971	548
1981	685

Source: India Economic Information Year Book, 1988-89.

SIMPLE BAR DIAGRAM

POPULATION IN INDIA (1901-1981)



Steps:

1. The years are represented on the horizontal line because the convention is to represent the independent variable along the horizontal line.

2. The relative population figures have been represented on the vertical line because the convention is to represent dependent vriable along the vertical line.

3. At each of the year i.e. 1901, 1911, 1921 etc. bars whose heights are proportional to the population figures are erected. The width of all these bars is the same.

Sometimes the bar diagrams are drawn with the bars horizontal instead of verticle. It does not make any difference if we decide to draw the horizontal bars or vertical bars.

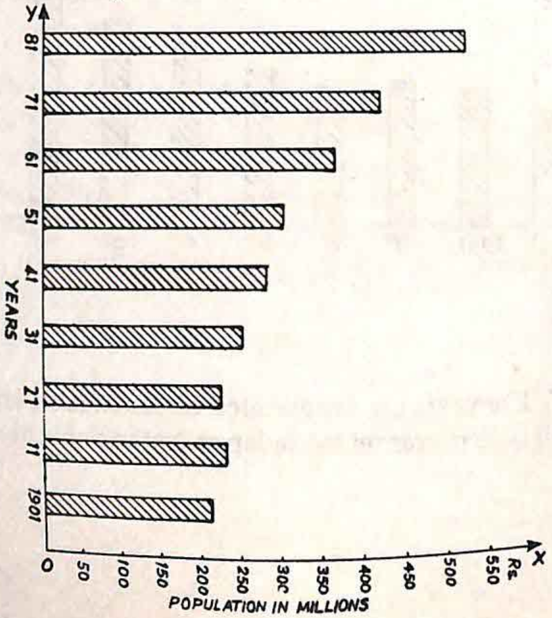
Illustration No. 2

The following table gives the figures of rural popoulation in India. Draw a bar diagram with the help of the table

Rural Population in India 1901-1981

Census year	Rural population(million)
1901	213
1911	226
1921	223
1931	245
1941	274
1951	298
1961	360
1971	439
1981	525

Source: Registrar General, India.



In drawing the above bar the same rules which were observed in drawing the vertical bar diagram except that the year of census of population along the vertical line and the corresponding population of rural area along the horizontal line.

(b) Sub-divided bar diagrams

These diagrams show the total of values and its break up into parts. *The bars are sub-divided into various parts in proportion to the values given in the data and a bar as such represents the total.* These diagrams are also called *Component Bar Diagrams*; we can use different colours like black, green, red or different shades of colour to distinguish sub-divisions of a bar. One should always remeber that the various components shoule be kept in the same order in each bar.

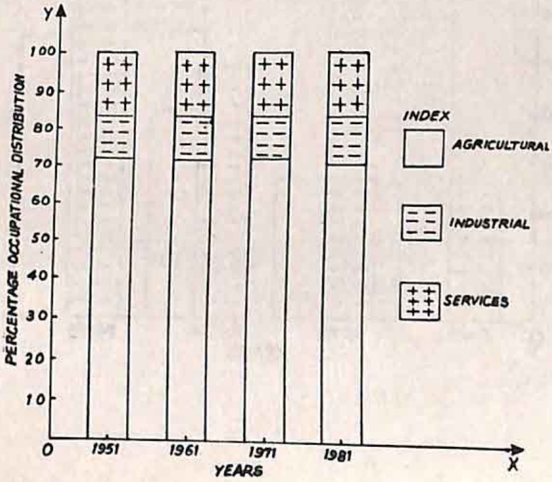
Illustration No.3

Represent the following data by means of a sub-divided bar diagram.

Percentage occupational Distribution of working population 1951-81.

Sector	1951	1961	1971	1981
Agricultural Sector	72.1	71.8	72.1	70.6
Industrial Sector	10.7	12.2	11.2	12.9
Services Sector	17.2	16.0	16.7	16.5
Total	100.00	100.00	100.00	100.00

Source: India Economic Information 1988- 89.



The sub-divided bar diagram or component part diagram is very useful to make comparison.

(C) *Multiple Bar Diagrams:* We use such diagrams when a comparison is to be made between two or more variables which are inter-related. The rules for drawing multiple bar diagrams are the same as the ones for simple bar diagrams. Separate bars are drawn for each value. We can compare one variable over diferent years and different variables at the same time giving different colours or shades to these variables.

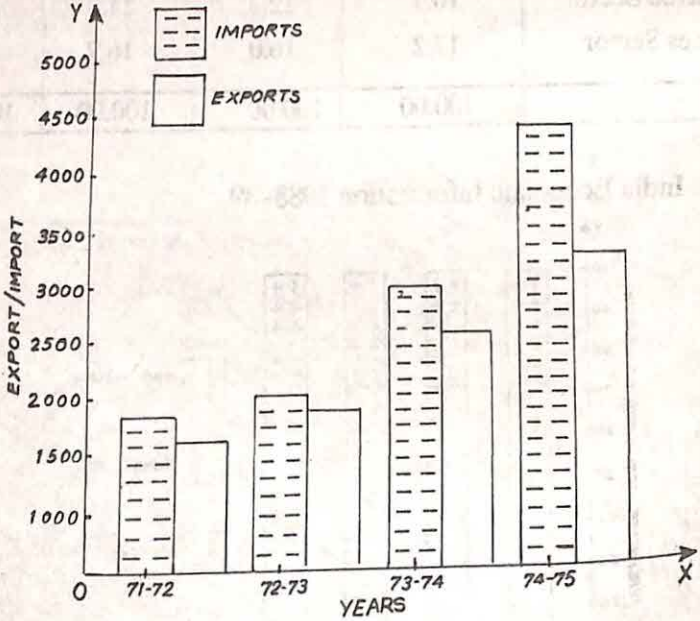
Illustration No.4

Present the data given below in the following table in Multiple bar diagram.

Year	Imports in crore rupees	Exports in crore rupees
1971-72	1825	1608
1972-73	1867	1971
1973-74	2925	2523
1974-75	4349	3253

MULTIPLE BAR DIAGRAM

Multiple Bar Diagram showing Exports and Imports.



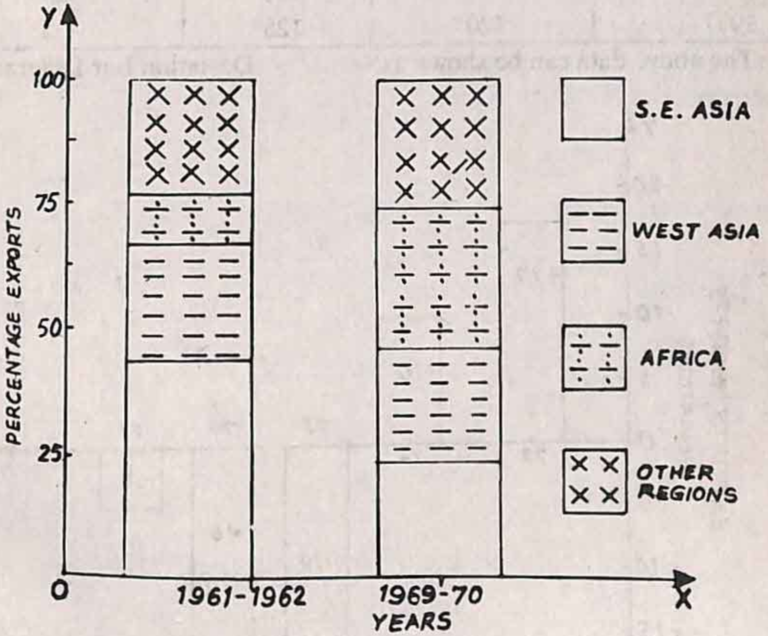
(D) Percentage bar diagram: When sub-divided bar diagram is drawn on percentage basis it is called percentage bar diagram. All the values are changed into percentage.

Illustration No.5

The percentage distributions of exports from India of engineering products by the region of destination for two years 1961-62 and 1969-70 are given below:

Region of destination	Year	
	1961-62	1969-70
S.E.Asia	42	23
West Asia	19	22
Africa	16	28
Other regions	23	27
	100	100

Draw a suitable diagram to represent the above data.



(E) Broken Bars: Sometimes we may get a series in which some values may be very small and others are very large. In this situation the largest bar or bars may be broken to gain space and to avoid the ugly look of the diagram.

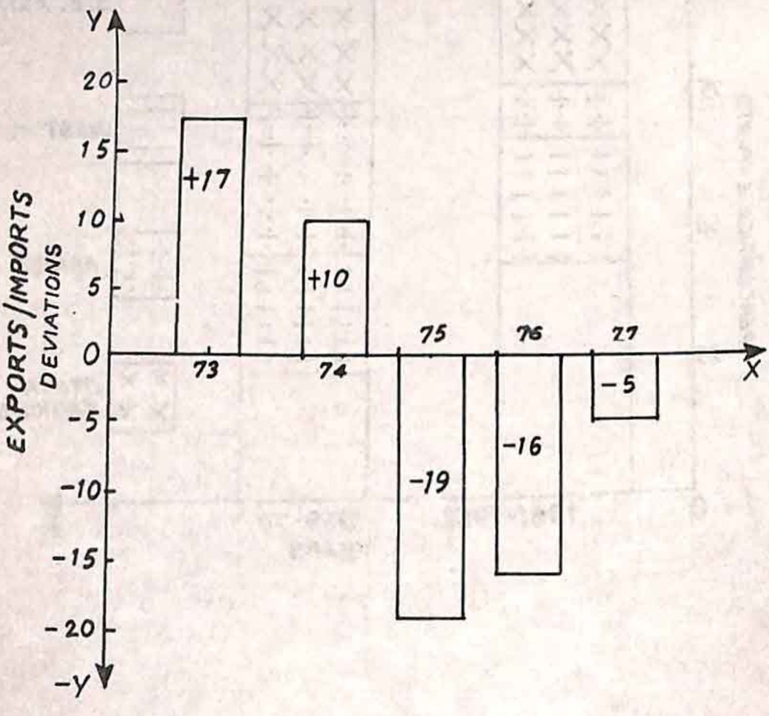
*(F) Deviation Bars:*This diagram is used to show plus and minus values in the same diagram. For example we are given the following figures of exports and imports and balance of trade in plus and minus.

Illustration No.6

Note: Hypothetical (in core rupees)

Year	Export	Import	Balance of Trade
1973	47	30	+17
1974	125	115	+10
1975	20	39	-19
1976	94	110	-16
1977	120	125	-5

The above data can be shown as:- Deviation Bar Diagram



Deviation Bars showing Balance of Trade

Two Dimensional Diagrams

If the statistical data consists of two variables then the data is represented with the help of two dimensional diagrams. In the case of one dimensional diagrams only the length is taken into account, the width is just for beautification. But in the case of two-dimensional diagrams we consider length as well as breadth. The most important types of such diagrams are:

(a) Rectangles

(b) Squares

(c) Circles or Pie-diagram.

(a) *Rectangles* :Rectangle diagram is used when two or more magnitudes with different components have to be compared. While drawing the rectangles, we take into account the length as well as the width of the bars. The area of the rectangle is equal to the product of the length and width of the bars. Rectangles help in representing statistical data in the original form or units as well as in rates or percentages. In this way a comparison between the various components becomes easy. Rectangles are kept side by side for comparison. This is a popular form of presentation of data.

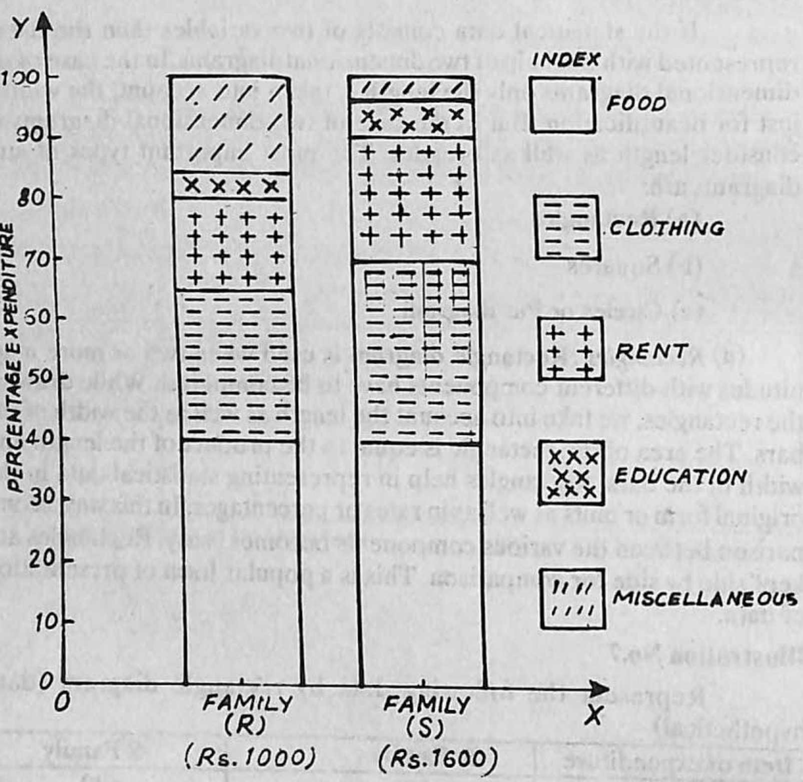
Illustration No.7

Represent the following data by rectangle diagram (data hypothetical)

Item of expenditure	R Family	S Family
Food	400	640
Clothing	250	480
Rent	150	320
Education	340	100
Miscellaneous	160	60
	1000	1600

Solution

Expenditure Items	Family R %	Family S %
Food	40	40
Clothing	25	30
Rent	15	20
Education	4	6.25
Miscellaneous	16	3.75
Total	100	100



(b) Squares: When the difference between the largest and smallest values of the variable representing a certain characteristic is very large it is not feasible to draw the bar diagram what so ever scale may be used. In such cases we represent the data with the help of squares or circles. The area of these squares or circles are proportionate to the values of the variable. Before drawing the circles or squares we find out the square roots of the values of the variable and divide these square roots by a suitable number to get the sides of the squares or radius of the circle. In order to have an attractive diagram and to facilitate comparison the squares are drawn with their sides on the same horizontal line.

Illustration No.8

Draw (i) square diagram and (ii) circle diagram to represent the following data given in the census figures.

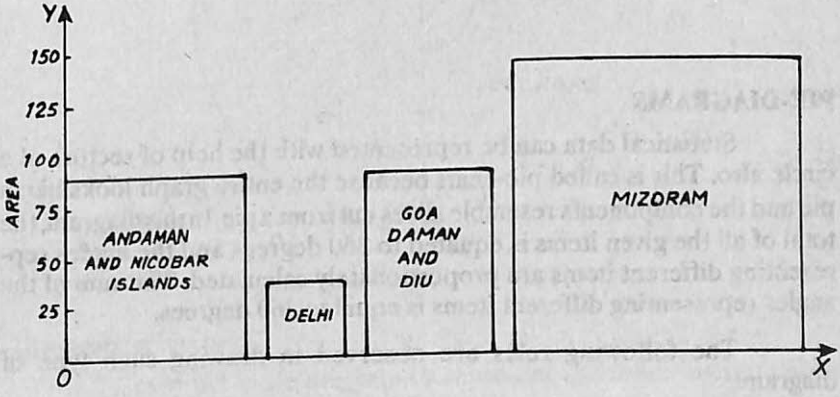
Name of the U.T. and States	Area in Sq. km.	Square Roots	side of square in cm.
Andaman & Nicobar Islands	8293	91.07	3.036
Delhi	1485	38.54	1.285
Goa, Daman andDiu	3813	61.75	2.058
Mizoram	21087	145.21	4.840

Solution
Steps

1. Take square roots of the area in sq. km.

2. Take a suitable common number of all the sq. roots and divide all the square roots by that number. Here we take 30 as a common number. Thus we get 1 cm = 30 km.

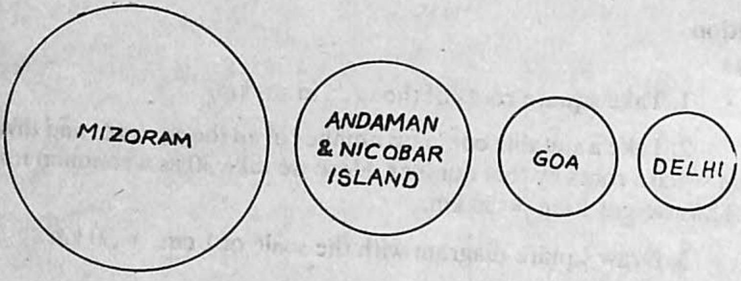
3. Draw square diagram with the scale of 1 cm. = 30 km.



(c) Circle Diagram

The area of a circle is πr^2. Therefore to draw the circle we take the radius of each circle proportional to the square root of the given magnitude. Thus, the radius of the circles will be the same as the side of the square diagram for the above data.

The circle diagrams are as under:



PIE-DIAGRAMS

Statistical data can be represented with the help of sectors of a circle also. This is called pie-chart because the entire graph looks like a pie and the components resemble slices cut from a pie. In this diagram, the total of all the given items is equated to 360 degrees and the angles representing different items are proportionately calculated. The sum of the angles representing different items is equal to 360 degrees.

The following rules are observed in drawing such type of diagrams:

(i) The total of all the given items is equated to 360 degrees and the angle representing each item is calculated because we know that the circle subtendes an angle of 360 degrees at the centre.

(ii) We divide the circle into different sectors each subtending an angle allotted for each item of the given data.

Illustration No.9

Construct a pie diagram with the help of the following information.

In a school 60% of the money spent goes for salaries, 25% for students welfare and 15% for building maintenance.

Solution

To construct pie diagram we should get the following calculations on the basis of 360 degree taken as equal to the total of the values.

Item	% of the money spent	Degree of angle
Salaries	60	216
Welfare	25	90
Building	15	54
	100	360

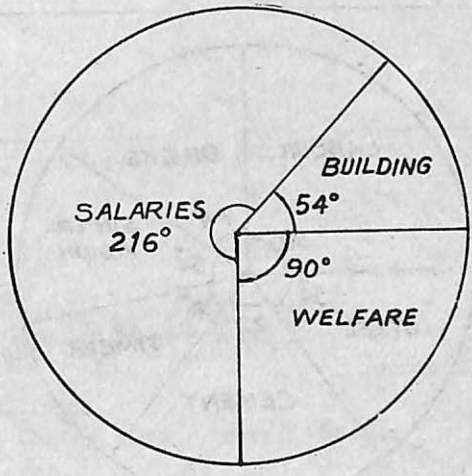


Illustration No.10

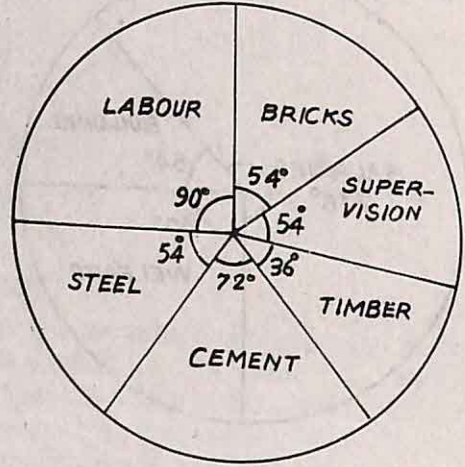
Construct a pie diagram to represent the following data about percentage break up of the cost of construction of a house in Delhi.

Item	Percent expenditure
Labour	25%
Bricks	15%
Cement	20%
Steel	15%
Timber	10%
Supervision	15%

Solution

To construct pie diagram we have to obtain the degree of angles by equating the total of values to 360 degree. In this case we get the following figures :

Items	% expenditure	Degree of angle
Labour	25	90
Bricks	15	54
Cement	20	72
Steel	15	54
Timber	10	36
Supervision	15	54
	100	360



PICTOGRAM

The technique of presenting data by pictograms was developed by Dr. Otto Neurth who was the resident of Vienna. This method is also known as the Vienna method or Isotype method. In recent years there has been an increasing use of pictograms in which the variable is represented by picture. Pictograms are easily understood and create a lasting impression on the mind. This method is particularly used for education, advertisement and propaganda purposes. Statistical data are represented through a pitcorial symbol carefully selected which should be self explanatory. As pictograms are more attractive, therefore, they are properly used where the attention of masses is to be drawn, such as exhibitions, fairs etc.

Pictograms are in fact highly simplified versions of diagrams. The pictorial symbol should be self-explanatory in nature and should represent the phenomena clearly. For example, the production of food should be shown with the help of ear-head *of wheat*. The increase in the number of literates should be represented with the help of pictures of men and women, boys and girls.

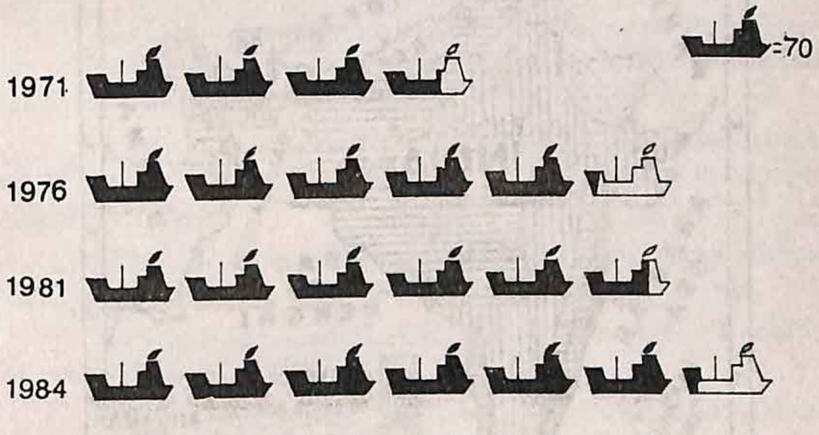
Illustration No.11

The following table gives the total number of vessels in 1971, 1976, 1981, 1984, and 1985 in Indian Shipping Fleet. Represent the data by means of a pictogram:

Year	No. of vessels
1971	255
1976	359
1981	403
1984	433
1985	488

Source: Ministry of Shipping and Transport.

Solution



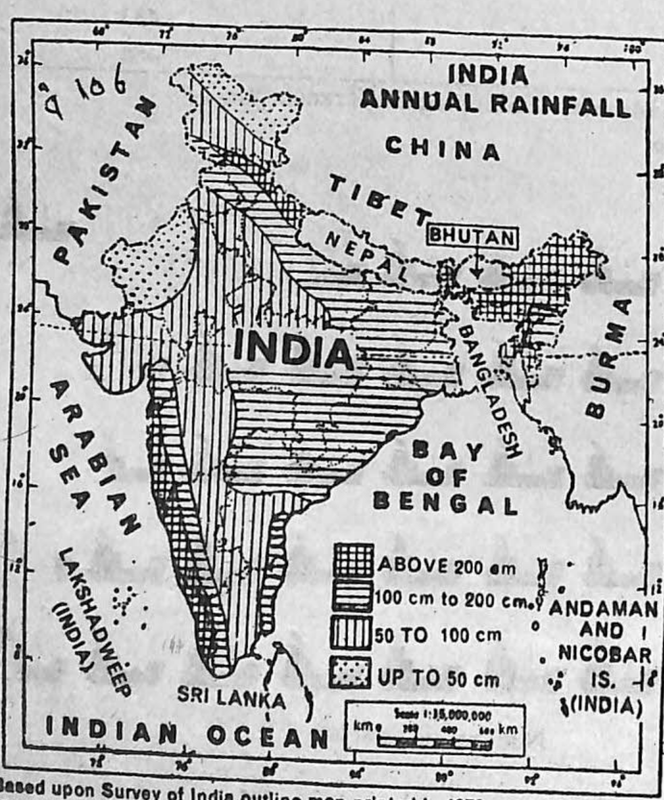
Number of vessels

CARTOGRAMS OR MAPS

Cartograms are the representation of statistical data with the help of maps. These are used to give quantitative information on a geographical basis. The differentiation between the variables or the numerical value of the variable on the map can be shown in many ways such as through different colours, shades, dots, symbolic representation in each geographical region or by placing the appropriate quantitative figures in each geographical region.

Cartograms should be used only when comparison is required on a geographical basis. Everyday on T.V. the Meteorological Department of Government of India explains the weather forecast and the temperature recorded in different parts of the country with the help of cartogram taken from Insat-I A

CARTOGRAM



Based upon Survey of India outline map printed in 1979.
The territorial waters of India extend into the sea to a distance of twelve nautical miles measured from the appropriate base line.
© Government of India copyright, 1981.

Choice of a suitable diagram

After studying the various types of diagrams now the question arises which diagram should be used for a particular data. There is no restriction what so ever. Following factors will decide about our choice of the diagram:

(a) The nature of the data.

(b) The available space where the diagram is to be drawn.

(c) The type of the population for whom the diagram is meant.

(d) The size of the data.

(e) The specific purpose of the investigation.

Generally the following points are kept in mind at the time of using particular type of diagram:

1. Bar diagram is useful when the deviations of the totals is to be studied or conveyed.

2. Sub-divided bar diagrams or component bar diagram could be used when the deviations of the totals as well as in the size of the components of the same characteristics are required to be studied or conveyed.

3. Multiple bar diagrams could be used when the deviations in the absolute values of the component quantitieshave to be stressed and the overall total is of no significance.

4. Square or circle diagrams could be used in those cases where the difference between the smallest and the largest values to be represented is very large.

5. Pie diagram could be used to show the relative proportions of the figures representing the portions of the same characteristics.

6.Deviation bars could be used to show the positive and negative aspects of the same attribute.

7. Pictograms could be used when the representation is meant for the layman.

8. Cartograms could be used when the statistical data is pertaining to some geographical units.

QUESTIONS

1. What is a pie digram ?

2. State the objectives of Tabular presentation of data.

(Hint: objectives of tabulation are : (a) Presenting data in an organised and scientific manner. (b) Presenting data in a condensed form for easy comprehension. (c) Simplification of problem (d) Facilitating comparison of data and (e) Saving time and space).

3. Present the data given below in the following table in multiple bar diagram:

Year	Imports	Exports
1971-72	1825	1608
1972-73	1867	1971
1973-74	2925	2523
1974-75	4349	3253

4. Write short notes on:

(a) Pictogram

(b) Pie-diagram

(c) Cartogram

5. What are the uses of diagrammatic presentations ? Name the different types of diagrams used in statistics.

6. Distinguish between the bar and pie diagrams.

7. What are the merits and limitations of diagrammatic representation of statistical data ?

8. What are the objectives of diagrams ?

9. Explain the various rules of drawing a diagram.

10. Write short notes on:

(a) Square diagram

(b) Rectangular diagram

(c) Multiple Bar Diagram.

(d) Sub-divided percentage bar diagram

11. What is tabulation ?

12. Describe the major functional parts of statistical tables. Draw a format of a table showing all these parts. (Hint main parts of a table:- (a) Table number, (b) Title of the table (c) Caption and stubs (d) Head notes (e) Body (f) foot notes and (g) source notes).

13. Write short note on the objectives of tabulation.

14. What points should be taken into account while preparing a table ?

15. What are the various types of tables used in an economic survey ?

16. Distinguish between tabulation and classification.

17. Discuss briefly the characteristics or essentials of a good table.

18. The percentage share of "Domestic Product" at factor cost by industry of origin is given below for 1960-61 and 1972-73. Prepare a pie diagram showing the changing share of different sectors .

Sectors	Years	
	1960-61	1972-73
Primary	52.2	41.7
Secondary	19.1	23.3
Teritiary	28.7	35.0
	100.00	100.00

19. The percentage distribution of exports from India of engineering products by the region of destination for two years 1961-62 and 1969-70 are given below:

Region of destination	YEAR	
	1961-62	1969-70
S.E.Asia	42	23
West Asia	19	22
Africa	16	28
Other regions	23	27
Total	100	100

Draw a suitable diagram to represent the above data.

20. Present the following data by sub-divided bar diagram:

IMPORTS (Rs. in crores)

Items	1972-73	1973-74	1974-75	1975-76
Food	128	474	795	988
Fertilisers	77	125	298	323
Mineral Oil	203	341	1113	1042
Others	1738	1789	1951	2394
Total	2146	2729	4157	4744

21. Prepare multiple bar diagram. Table showing imports and exports

Year	Imports	Exports
1973-74	2955	2523
1974-75	4519	3329
1975-76	5265	4049
1976-77	5265	5143

22. Draw a simple bar diagram
Production of coal (million tonnes)

Year	Production
1983-84	138.2
1984-85	147.4
1985-86	154.2
1986-87	165.8

23. Monthly incomes of five families A, B, C, D and E are Rs. 3600/-, Rs 1600/-, Rs.900/-, Rs.100/, and Rs.75/-. Represent this data by square and circle diagrams.

24. Represent the following data with the help of a pie-diagram:

Items of expenditure	Expenditure	
	A family	B family
Food	430	830
Clothing	80	170
Recreation	30	100
Education	50	90
Rent	100	210
Miscellaneous	60	150

25. Construct a pie diagram to represent the following data about the percentage breakup of the cost of construction of a house in Delhi.

Item Percent	Expenditure
Labour	25
Bricks	15
Cement	20
Steel	15
Timber	10
Supervision	15

2

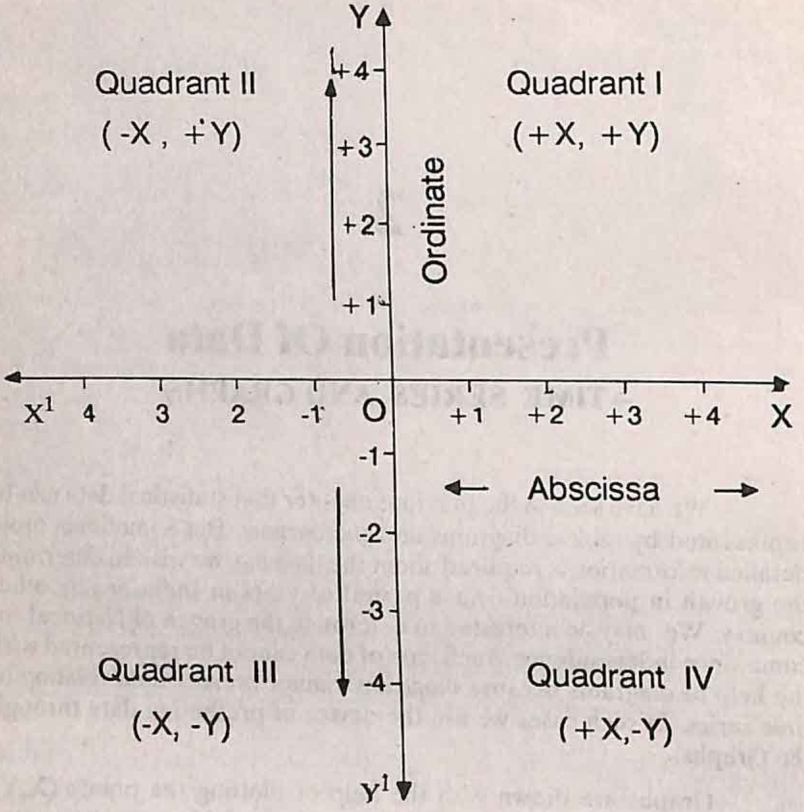
Presentation Of Data

– TIME SERIES AND GRAPHS

We have seen in the previous chapter that statistical data can be represented by tables, diagrams and pictograms. But sometimes more detailed information is required about the data e.g. we wish to determine the growth in population over a period of years in India or any other country. We may be interested to determine the growth of National Income since independence. Such type of data cannot be represented with the help of diagrams because diagrams cannot present data relating to time series. In such cases we use the device of presenting data through the Graphs.

Graphs are drawn with the help of plotting the points (X,Y) where "X" denotes the value of an independent variable and "Y" denotes the value of a dependent variable. These plotted points are joined with the help of line or curves to get the required graph. Plotting of these points is done with the help of coordinates. These coordinates are obtained with the help of two straight lines intersecting each other and drawn at right angles to each other in a plane and with the help of these intersecting lines it is possible to describe the position of any point in the plane with reference to the point of intersection of the two lines. These two straight lines are called as X and Y axis. The point of intersecting is called the point of origin or "O" which has zero value and the horizontal line is called theX-axis and the vertical line Y axis. Distances measured towards the right and upward from the origin are "positive" and those measured towards the left and downward are negative.

It is customary to call verticle line as Y-axis (Ordinate) or YY^1 and horizontal line as x-axis (Abscissa) shown as XX^1. Both the lines are known as coordinates axis. This is depicted in the figure given on the next page.



The plane paper or graph paper is divided into four equal parts called quadrants.

Quadrant I. As shown in YOX, represents positive values of X and Y.

Quadrant II. Area in YOX^1, where Y is positive and X is negative.

Quadrant III. Area in $X^1 OY^1$ where both the values of X and Y are negative.

Quadrant IV. Area in XOY^1 where X is positive and Y is negative.

Generally in economic analysis we use Quadrant I i.e. YOX. Sometimes Quadrant IV is also used where X is positive and Y is negative. In economic analysis the statistical data is expressed often in terms of two or more than two related variables-one of which is independent and the other dependent. We represent the independent variable on the X-axis and the dependent variables on the Y-axis.

Rules for drawing Graphs

The following rules are observed in drawing the graphs:-

(a) *Heading*: Every graph must have a heading title which should convey the sum and substance of the graph.

(b) *Choice of scale*: The scale should be chosen according to the size of the numerical data and the size of the graph paper available.

(c) *Details of scale*: The details of the scale used in plotting the graph must be mentioned in the graph.

(d) *Propotion of Axes:* As far as possible the ratio between scales on X-axis and Y-axis should be 3:2.

(e) *Use of false base line:* In case the smallest value of the dependent variable is at a great distance from the origin, we break the verticle scale near the origin and take any convenient point on the vertical line to represent the smallest value of the dependent variable and also join that point to the origin with the help of zigzag line. This techinque of selecting the point on the Y-axis is called false base line method.

(f) *Sketching of a curve:* After plotting the points on the graph paper we should join them with the help of either straight lines or curves. These straight lines/curves should be of same thickness in the same graph.

(g) *Diversification of curves:* If more than one graph is drawn on the same horizontal scale then the points of the different graphs should be joined with the help of different straight lines/curves and their separate identity must be maintained.

(h) *Index or Indicator of graphs*: In order to differentiate between the different graphs drawn on the same base, these graphs must be indexed.

USE OF GRAPHIC METHOD

The graphic method is generally used in the following two cases:

(i) To represent time-series data and

(ii) To represent data in a frequency distribution.

TIME SERIES GRAPH

The information arranged over a period of time e.g. years, months, weeks, days etc. is termed as a time series. Time series can be shown on the graph paper. Presentation of this type of information by line or curve on the graph paper is of great use in economic statistics. The time period is always represented on the X-axis and the corresponding values of the other variable/variables along the Y-axis.

Illustration No.1

Time series with single dependent variable

From the following table prepare a graph.

Years	No. of students in a school
1981	50
1982	150
1983	100
1984	150
1985	200
1986	225
1987	200
1988	250

GRAPH SHOWING NUMBER OF STUDENTS IN A SCHOOL

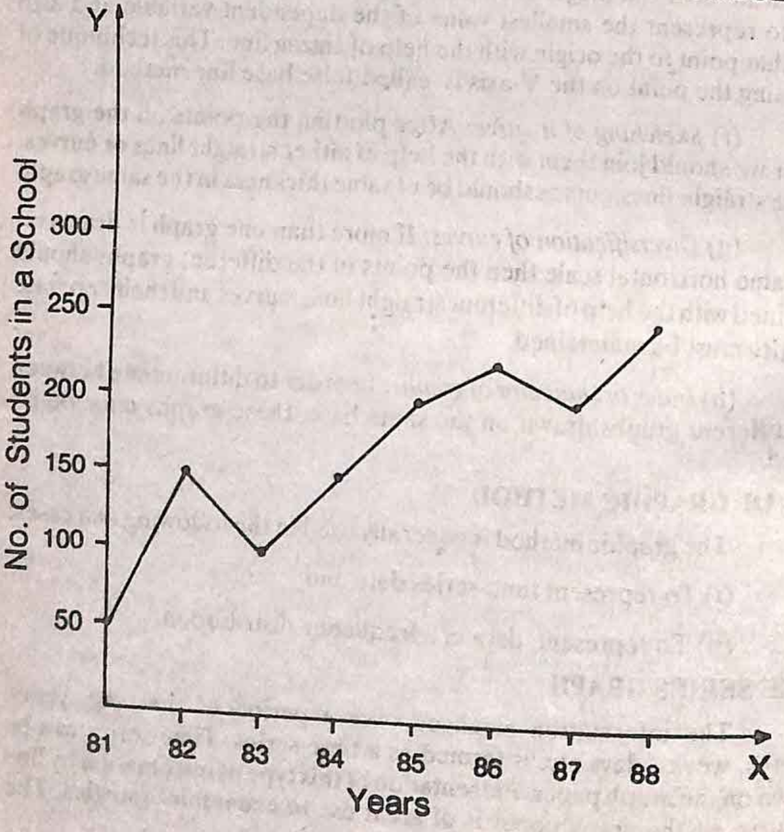


Illustration No.2

Present the following information graphically.

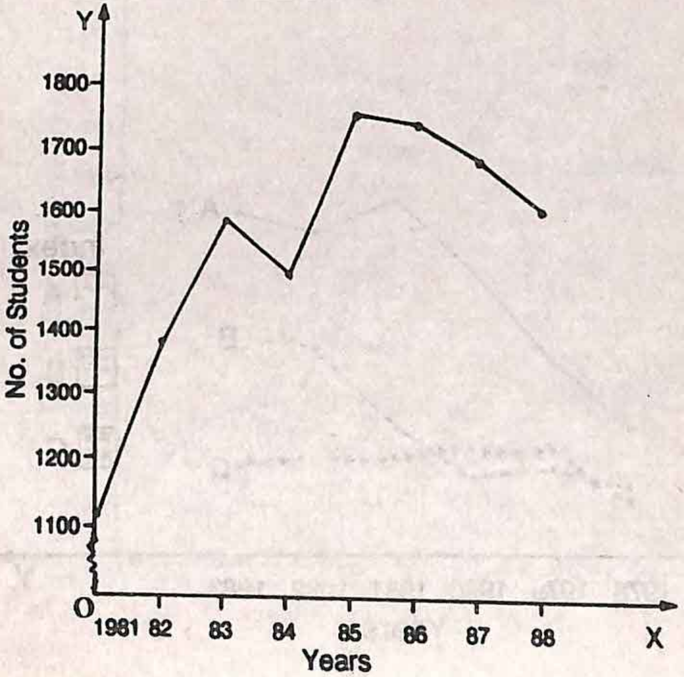
Central School, New Delhi.

Years	Students
1981	1120
1982	1380
1983	1587
1984	1490
1985	1760
1986	1734
1987	1675
1988	1600

Solution: The number of students is more than one thousand in each year but changes in magnitudes from one year to another are in hundreds. If we start Y-axis having the scale with the difference of say 100 or 200 or 500, we might be using small space to get a line graph and wasting too much of space of graph paper. So we use false base line in graphic presentation.

Thus by using false base line we represent the above data through graph as under:-

GRAPH WITH FALSE BASE LINE SHOWING NO. OF STUDENTS



GRAPHS OF TWO OR MORE THAN TWO VARIABLES

We can show on the graph paper two or more than two variables provided that all the variables are in the same unit of measurement. One common scale for Y-axis is prepared to plot the given values. It is advise-able to use different types of lines whenever two or more values are shown on the same graph e.g. a dotted line, a broken line or a thick line etc. One graph may be used to plot a maximum of 5 to 6 variables at a time. The index is to be shown on the graph paper.

Illustration No.3

Depict the following data with the help of a graph:

Annual Circulation (in crores)

	1978	1979	1980	1981	1982	1983
Daily A	15	22	35	44	40	43
" B	7	10	9	15	25	26
" C	6	12	12	13	12	10

ANNUAL CIRCULATION OF THREE NEWSPAPERS

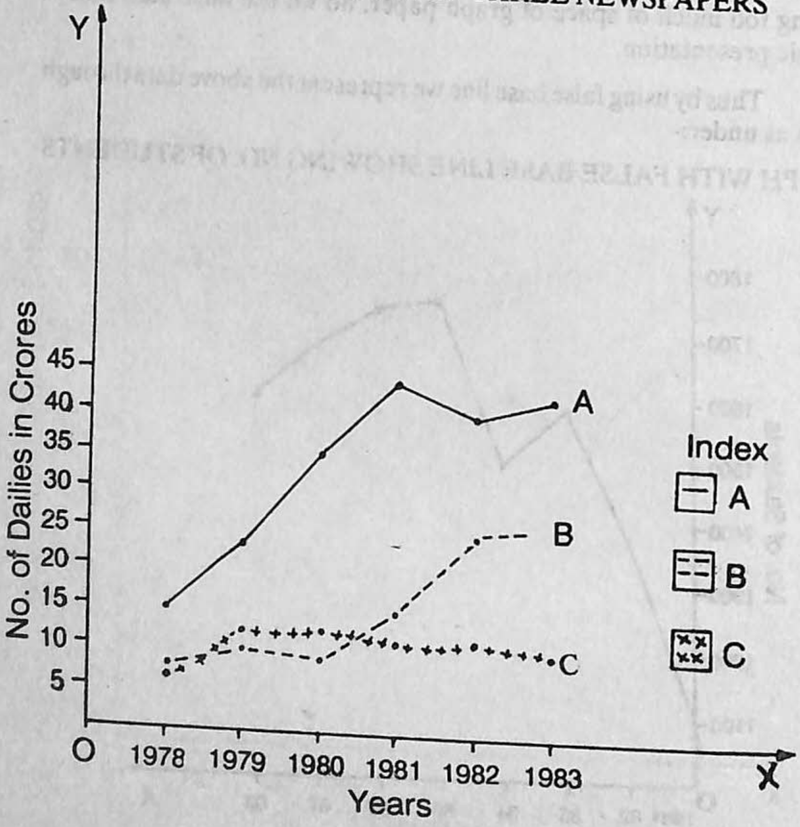


Illustration No.4

Prepare a graph representing the following data of consumer's price Index.

	1970-71	1971-72	1972-73	1973-74
Delhi	199	217	224	265
Calcutta	182	190	204	236
Bombay	182	188	196	228
	1974-75	1975-76	1976-77	1977-78
Delhi	338	334	332	358
Calcutta	287	300	297	318
Bombay	284	285	298	320
	1978-79			
Delhi	368			
Calcutta	325			
Bombay	231			

GRAPH SHOWING COST OF LIVING INDEX

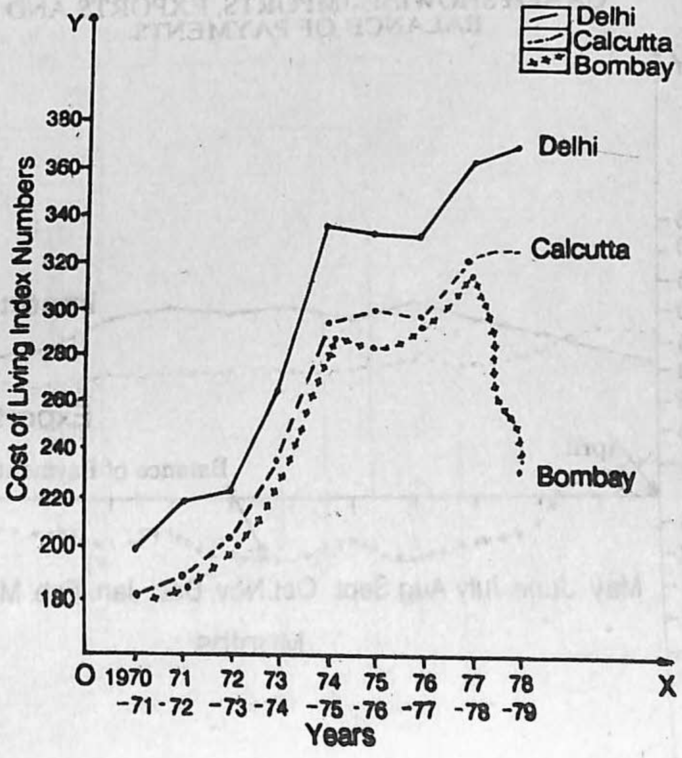
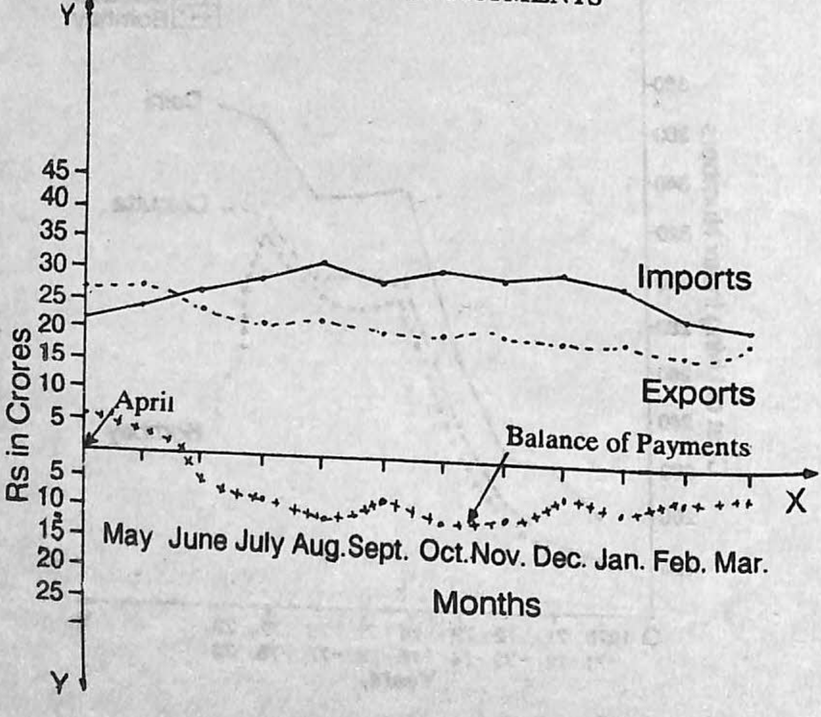


Illustration No.5

From the following table of Imports and Exports prepare a time series graph showing all the variables.

Month	Imports	Exports	Balance
April	22	28	+6
May	24	28	+4
June	26	23	-3
July	28	21	-7
August	31	20	-11
September	29	22	-7
October	32	21	-11
November	30	20	-10
December	32	20	-12
January	31	20	-11
February	25	18	-7
March	24	20	-4

GRAPH SHOWING IMPORTS, EXPORTS AND BALANCE OF PAYMENTS



Graphs of Frequency Distribution

When the data is classified as frequency distribution the graph representing this data is called frequency graph. While plotting the frequency graph the values of the variable are represented on the X-axis and their corresponding frequencies on the Y-axis. There are five types of frequency graphs used in statistics.

They are:-

(i) Line Frequency diagram.

(ii) Histogram.

(iii) Frequency polygon.

(iv) Frequency curve.

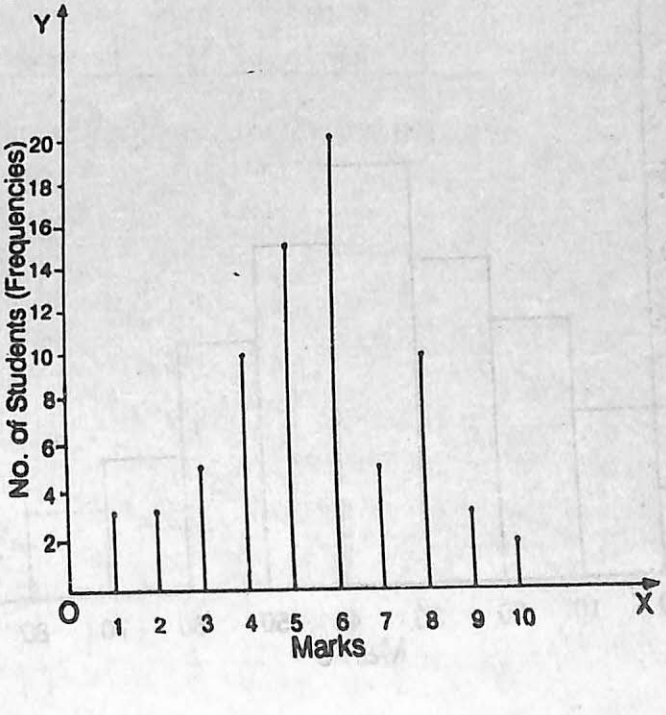
(v) Cumulative Frequency curve or Ogive.

(i) *Line Frequency Diagram:* For discrete distribution we prepare Line Frequency Graph or Frequency Bar or Line Frequency Diagram. The size of the item is taken in ascending order of magnitude and represented along the X-axis and the frequency is represented along the Y-axis. On the value of the variable a line with a weight equal to the number of its frequency is drawn.

Illustration No.6

Draw a Line frequency Graph of the following data:-

Marks	1	2	3	4	5	6	7	8	9	10
No. of students (frequency)	3	3	5	10	15	20	5	10	3	2



Steps:

1. Marks taken on X-axis.
2. No. of students (frequencies) on Y-axis.
3. Draw a vertical line on each value equal to the length of each frequency.

(ii) *Histogram:* It is a two dimensional diagram. Histogram is used to represent graphically the *grouped frequency distribution data*. In drawing the histogram we first mark along the X-axis all the class intervals on a suitable scale. With the class-intervals as bases, we draw rectangles whose areas are proportional to the frequencies of the class-intervals.

There are two types of histograms: one for the equal class-intervals and the other for the unequal class-intervals. For equal class-intervals the heights of the rectangles will be proportional to the frequencies and for unequal class-intervals the heights of the rectangles will be proportional to the ratios of the frequencies to the width of the classes.

Histogram of Equal class intervals:

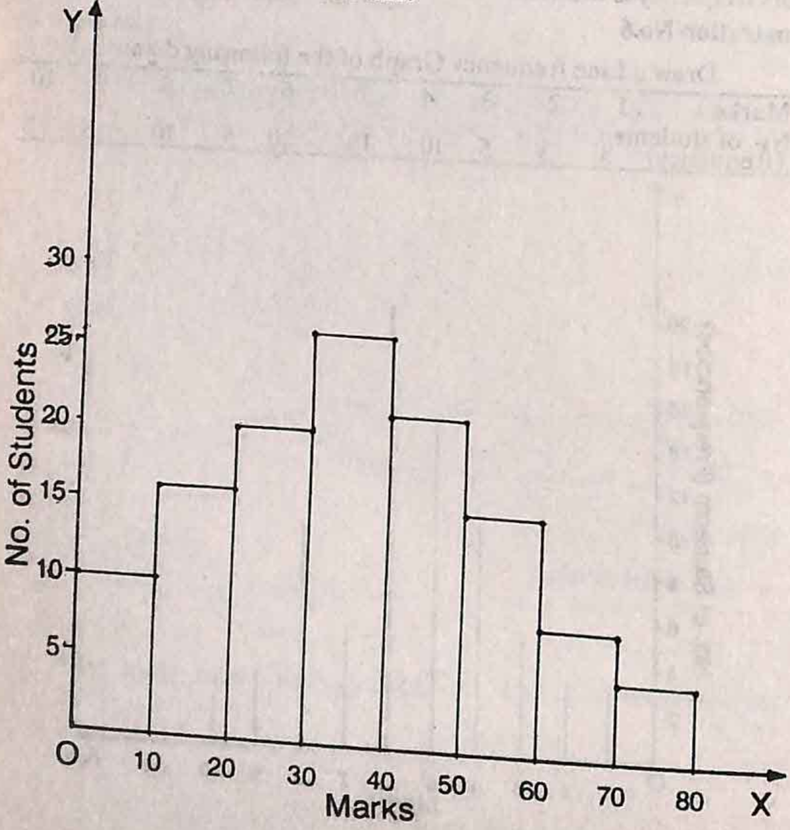


Illustration No.7

Draw a Histogram from the following data:

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	10	16	20	26	22
Marks	50-60	60-70	70-80		
No. of students	15	8	5		

Steps:

1. On X axis we take marks.
2. On Y axis we take frequencies
3. Draw rectangles joining each other representing each class with frequency.

Histogram is Shown On Page 119

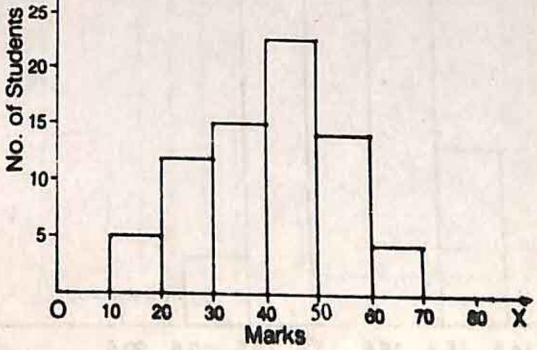
Illustration No.8

Make a Histogram using the data given below:-

Marks obtained	10-20	20-30	30-40	40-50
No. of students	5	12	15	22
Marks obtained	50-60	60-70		
No. of students	14	4		

Solution

HISTOGRAM OF EQUAL CLASS INTERVAL



Histogram when Mid-points are given

Illustration No.9

Construct a Histogram from the following distribution of total marks obtained by 50 students of a school in C.B.S.E.Examination.

Marks Mid-points	No. of students
150	5
160	10
170	20
180	12
190	2
200	1
	50

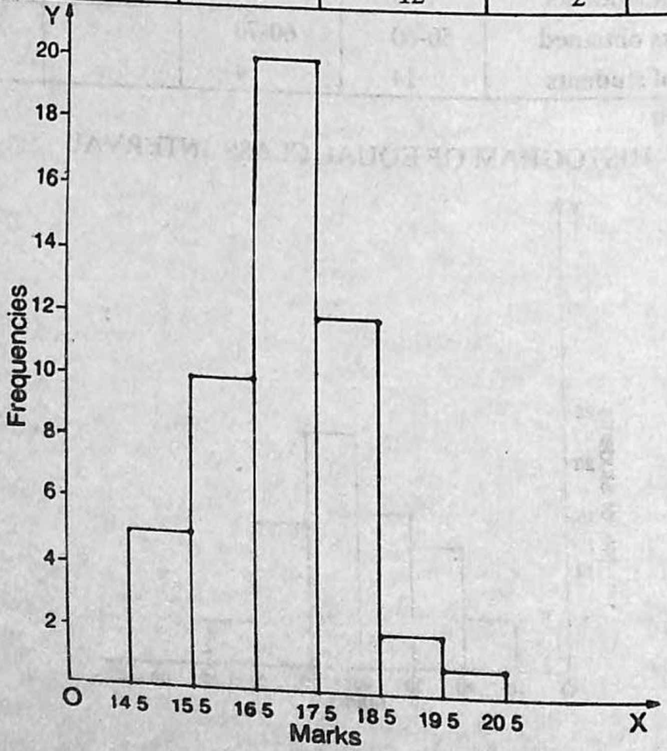
Solution

Steps:

1. First find out the class intervals from the midpoints.
2. On X-axis take the marks.
3. On Y-axis take the frequencies.
4. Draw rectangles joining each other representing each class with frequency.

Class intervals will be

145-155	155-165	165-175	175-185	185-195	195-205
5	10	20	12	2	1



Histogram of the unequal class intervals

Illustration No.10

Represent the following data by means of Histogram.

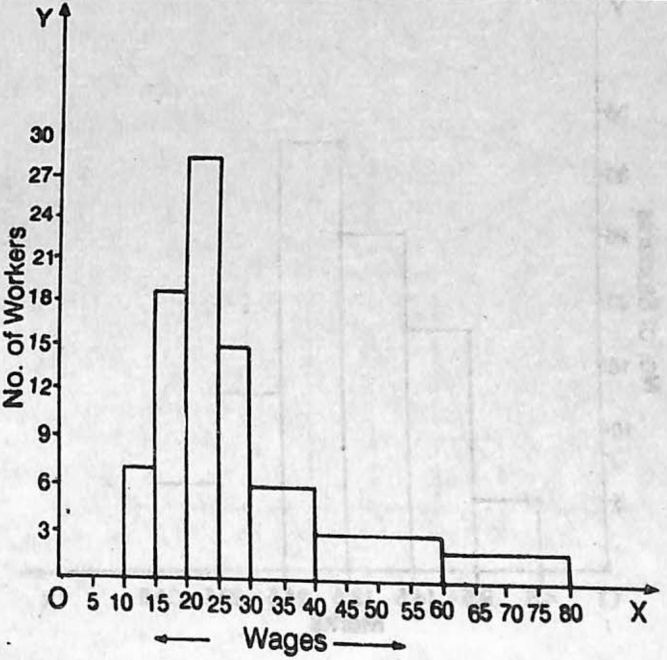
Wages	10-15	15-20	20-25	25-30	30-40	40-60	60-80
No. of Workers	7	19	28	15	12	12	8

Solution:

In order to construct histogram with unequal class intervals we first have to adjust the frequencies according to the frequencies of the lowest class. Here 5 is the lowest class interval. Frequencies of other classes will be adjusted with this lowest class interval in the following way.

Adjusted frequencies:

Wages	Frequency	Adjustment	New frequency
10-15	7		7
15-20	19		19
20-25	28		28
25-30	15		15
30-40	12	$\frac{5 \times 12}{10}$	6
40-60	12	$\frac{5 \times 12}{20}$	3
60-80	8	$\frac{5 \times 8}{20}$	2



Histogram when class intervals are given by Inclusive Method

Illustration No.11

Construct a Histogram from the following data:

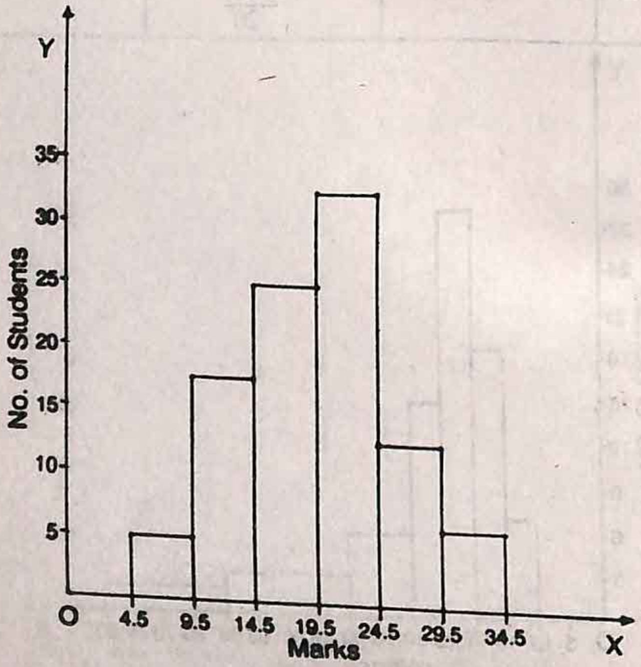
Marks	5-9	10-14	15-19	20-24
Students	4	17	25	32
Marks	25-29	30-35		
Students	13	6		

Solution

In this case when class intervals are given by inclusive method, it is necessary to convert them into exclusive class intervals. The adjusted class intervals will be:

4.5 - 9.5	4
9.5 - 14.5	17
14.5 - 19.5	25
19.5 - 24.5	32
24.5 - 29.5	13
29.5 - 34.5	6

HISTOGRAM WITH ADJUSTED CLASS INTERVAL



(iii) Frequency Polygon

Frequency Polygon can be drawn by two methods:

(a) With Histogram and

(b) Without Histogram.

Frequency Polygon with Histogram

For a grouped frequency distribution the *Frequency Polygon* can be obtained by joining the middle points of the upper side of the adjacent rectangles of the Histogram by means of straight lines.

Illustration No.12

Make a frequency Polygon and Histogram using the given data

Marks obtained	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	5	12	15	22	14	4

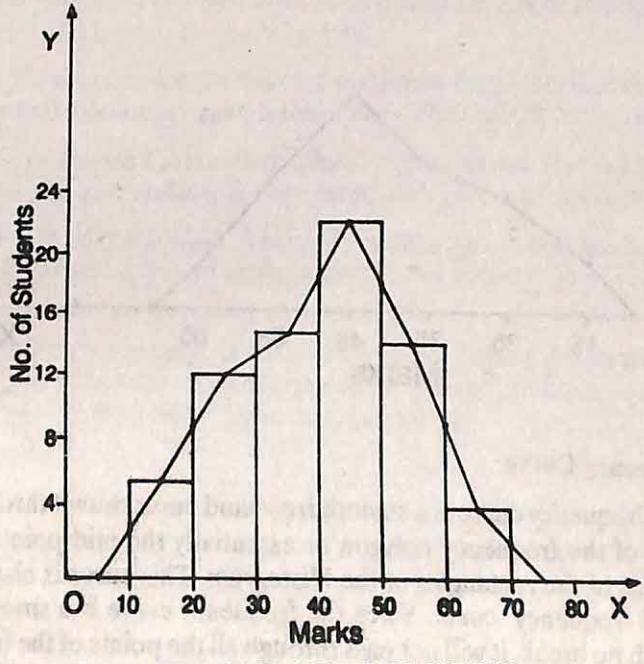
Solution

1. Draw a Histogram with the given data.

2. Obtain the mid points of the upper horizontal side of each rectangle.

3. Join these mid-points of the adjacent rectangles of the Histogram by straight lines.

4. Extend the ends of frequency polygon to the base line.



FREQUENCY POLYGON

Frequency Polygon without Histogram

Illustration No.13

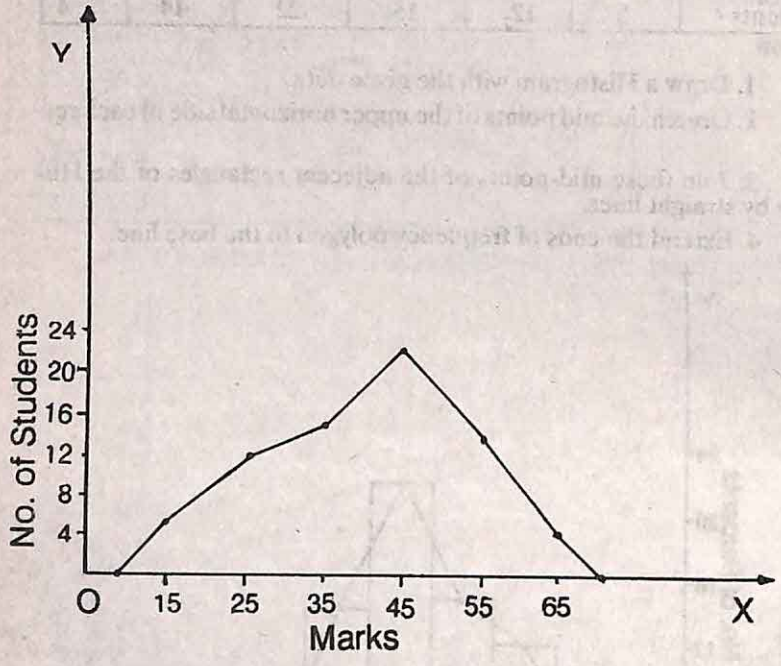
From the following data prepare frequency polygon.

Midpoints	15	25	35	45	55	65
No. of students	5	12	15	22	14	4

Solution

Steps:

1. Take the mid-points on X-axis
2. Take frequency on Y axis.
3. Join the points plotted for the mid-points corresponding to their frequencies by straight lines.



(iv) Frequency Curve

Frequency curve is a smooth free hand curve drawn through the vertices of the frequency polygon or essentially the mid-points of the upper sides of the rectangles of the Histogram. This curve is also called smoothed frequency curve. Since the frequency curve is a smooth one drawn with no break, it will not pass through all the points of the frequency polygon exactly but it will go through them as closely as possible, and so the points of the polygon will be clustered around the frequency curve.

Points to be kept in view while constructing frequency curve:

1. One should draw first frequency polygon to get a smoothed frequency curve.

2. Smooth free hand curve should not have any break in between.

3. For getting correct results smoothing the frequency polygon should be done with care.

4. If frequency curve is drawn with care the area of the frequency curve is approximately the same as the area of frequency polygon.

5. The total area under the frequency curve will be proportional to the total frequency. As there are different types of frequency distribution, similarly frequency curves are of various types giving different shapes derived from histogram and frequency polygon. Generally the curves are of following kinds:-

(a) *Normal curve:*It is also called bell shaped or symmetrical curve. Highest frequency is at the centre and both the tails on left and right hand sides are in identical fashion.

(b) *Positively skewed curve:*It is the asymmetrical (Positive) curve. It is having a tail towards right hand side of the graph and frequencies are more for lower values of the data.

(c) *U-shaped curve:*In this case maximum frequency is at the ends of the data and minimum towards the centre. The highest are at the end.

(d) *G-shaped Curve:*When class trequencies are running up to a maximum at one end and not at the centre, they give a "g" shaped curve.

(e) *Mixed curve:*When frequency of the given values are increasing and decreasing again and again, it provides a mixed type of curve.

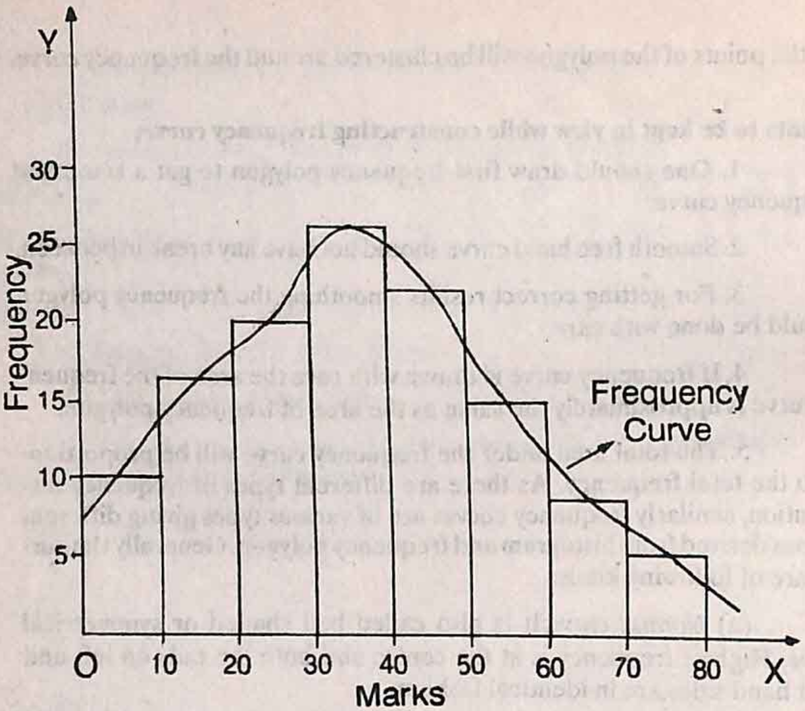
Method of Drawing a Frequency Curve:

The frequency curve is a smooth free hand curve drawn through the mid-points of the upper side of the rectangles of the histogram. We illustrate this with the help of following illustration:-

Illustration No.14

Draw Histogram and Frequency curve from the following data:

Marks obtained	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequencies	10	16	20	26	22	15	8	5



(v) Cumulative Frequency Curve

"Ogive" or "Cumulative frequency curve" can be obtained on the graph paper by two methods:-

(a) "Less than" method and

(b) "More than" method.

In "less than method" we plot the points as X and Y axises. The X axis represents the upper limit of the class interval and Y-axis represents the cumulative frequency of the respective class. These points are joined with the help of smooth curves and the resulting curve is the Ogive.

In the "more than" method X-axis represents the lower limit of the class interval and Y represents the cumulative frequency of the respective class in the descending order. These plotted points are joined with the help of smooth curves.

Illustration No.15:

Draw "Ogive" curves of 'Less than" and "More than" from the following data:-

0-10	10-20	20-30	30-40	40-50	50-60	60-70
4	4	7	10	12	8	5

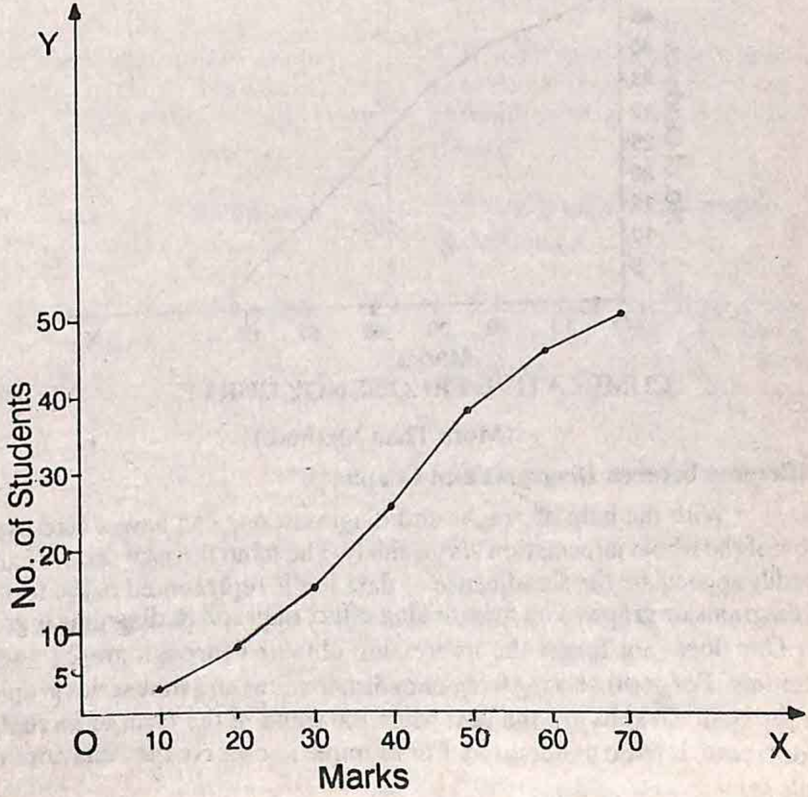
Solution

In "less than Method" we start with the upper limits of the classes and proceeding further by adding frequencies in the following way:

Cumulative Frequency Distribution By "Less than" method

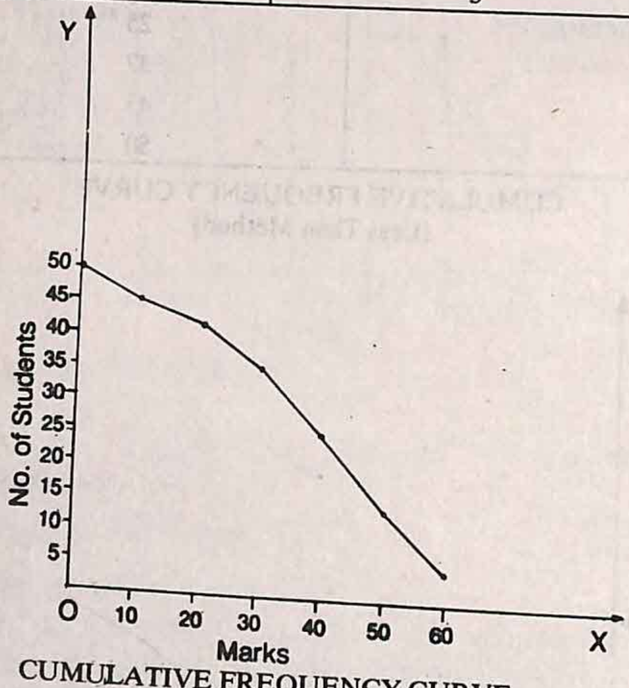
Marks Less than	No. of students
10	4
20	8
30	15
40	25
50	37
60	45
70	50

CUMULATIVE FREQUENCY CURVE
(Less Than Method)



Cumulative Frequency Distribution by "More than method."

Marks More than	No. of students
0	50
10	46
20	42
30	35
40	25
50	13
60	5



CUMULATIVE FREQUENCY CURVE

(More Than Method.)

Difference between Diagrams and Graphs

With the help of graphs and diagrams, one can have a bird eye view of the whole information very quickly. The mind through the eye can readily appreciate the significance of data if it is represented in the form of diagrams or graphs.The memorising effect of graph & diagrams is greater.One does not forget the impression obtained through graphs and diagrams. For representing frequency distributions and time series graphs are the best. Graphs are the best when the trend in the form of increase or decrease is to be understood. For example to observe the behaviour of whole sale-price index, graphs are the best for presentation of information. Graphs and diagrams help us in making quick comparisons.

DIFFERENCE BETWEEN

Diagrammatic Presentation	Graphic Presentation
1. An ordinary paper can be used in its presentation.	1. Graphs are constructed on a graph paper with the help of coordinates.
2. Are more attractive and suitable for publicity and programmes.	2. Are suitable for time series and frequency distribution.
3. Help us in comparison of the data and not suitable for further statistical treatment.	3. Are very much used by the statisticians and research workers and are suitable for further statistical treatment. Helpful to findout mode and various partition values.
4. More appropriate when data is expressed in geographic location.	4. It is more appropriate when the data is expressed in chronological or frequency distribution.
5. Can be understood even by a layman.	5. It needs expert knowledge to understand it.
6. It is effective when one or two variables are there.	6. It becomes operative when more than one series are to be presented.

QUESTIONS

1. Distinguish between graphic presentation and diagrammatic presentation of data. Explain with the help of examples.

2. What is "Cumulative Frequency Curve". Draw "Less than" and "More than" Ogive from the following data and indicate the value of the median:-

Marks	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
No. of students: (Ans. M = 22)	7	10	20	13	12	10	14	9

3. Distinguish between "Frequency Polygon" and "Histogram". Make a frequency polygon and Histogram using the given data:-

Marks obtained:	10-20	20-30	30-40	40-50	50-60	60-70
No. of students:	5	12	15	22	14	4

4. Name the commonly used graphs. Explain how Time series graph is prepared?

(Hint: Types of graphs-Line frequency graphs, Histogram, frequency polygon, freuency curve, Cumulative frequency curve or Ogive.Time series graphs are used to represent one or more than one variables. While explaining various components of one item we make use of zone graph. In a time series Graph on the X-axis time is depicted and on Y axis values are depicted).

5. Depict the following figures in a Time series graph.

Year	Profit in thousand rupees
1973	60
1974	72
1975	75
1976	65
1977	80
1978	95

6. Define Histogram, frequency polygon and frequency curve and draw these from the rollowing data:-

Marks obtained	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	10	16	20	26	22	15	8	5

7. Arrange the following information on a time series graph:-

Net Domestic product of India (at 1970-71 prices)

Year	(Rs. thousand crores)
1973-74	35
1974-75	36
1975-76	37
1976-77	40
1977-78	41
1978-79	44
1979-80	44

8. Explain the difference between frequency polygon and Histogram, giving illustrations.

9. Draw a Histogram:-

Age	16-19	20-29	30-39	40-49	50-59	60-64
Persons	15	46	49	32	28	14

10. In a certain Colony a sample of 40 households was selected. The data on monthly income for this sample are as follows:-

200, 120, 350, 550, 400, 140, 350, 85, 180, 110, 110, 600, 350, 500, 450, 200, 170, 90, 170, 800; 190, 700, 630, 170, 210, 185, 250, 120, 180, 350, 110, 250, 430, 140, 300, 400, 200, 400, 210, 300

(a) Construct a histogram and a frequency polygon.

(b) Show that the area under the polygon is equal to the area under the histogram.

{Hint:- Take a class interval of 100 starting with 0-100 and so on. Prepare frequency distribution. Then prepare Histogram and Frequency polygon. The sum of the area of the rectangles, forming the histogram is the same as that of the polygon becuase the area of the consecutive triangles (one left out from the histogram and the other included in the polygon not lying in the histogram)like- lined and dotted triangles are equal and similarly the areas of the other consecutive triangles)

angles)

11. (a) List the steps involved in constructing the histogram from data presented in the form of a frequency table. Explain how the construction procedure is simplified when all the class intervals have the same width.

(b) For a particular set of data on individual incomes, the "less than type" Ogive and the "more than type" Ogive were found to intersect at Rs. 715. Discuss the significance of Rs. 715/- for the given set of data.

[Hint: If all the class intervals are of the same size, then the length of the adjacent bar depends upon the class-frequencies. In such case no adjustment in frequencies for the purpose of histogramis requiredThe value at the joint of intersection is median. Rs.715/- is the median value.]

12. Write short notes on:

(a) Frequency polygon.

(b) Ogive

(c) Histogram

13. What is meant by Ogive. From the following distribution prepare the "less than" Ogive.

Capital (in lakh Rs)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Companies	2	3	7	11	15	7	2	3

14. The frequency distribution of marks obtained by students in a class test is given below.

Marks	0-10	10-20	20-30	30-40	40-50
No. of students:	3	10	14	10	3

Draw a histogram to represent the frequency distribution of marks. Comment on the shapes of the diagram.

(Hint: The Histogram shows that the distribution is symmetrical. This means that mean, median and mode are the same).

15. Prepare histogram from the following data:-

C.I.	10-15	15-20	20-25	25-30	30-35	35-40
Frequencies	3	5	9	7	4	2

16. Explain the various methods of presenting data graphically, giving suitable illustrations.

17. We have the following data on the monthly expenditure on food (in rupees) for 30 households in a locality:

115, 159, 196, 205, 112, 223, 256, 271, 310, 129, 135, 169, 184, 234, 245, 241, 265, 298, 144, 135, 172, 173, 229, 243, 220, 238, 278, 243, 220, 238

(a) Obtain a frequency distribution using class intervals 100-150, 150-200 and so on.

(b) Draw a frequency polygon.

(c) What percent spend less than 250/- per month and what per cent spend more than Rs.200/- per month ?

18. Plot the following data of annual profit of a firm on a time series graph:

Year	Profit in thousand rupees
1973	60
1974	72
1975	75
1976	65
1977	80
1978	95

19. What is meant by Histogram ?

Present the data given in the table below in the form of Histogram.

Midpoints	Frequency
115	6
125	25
135	48
145	72
155	116
165	60
175	38
185	22
195	3

(Hint: Convert the mid points into continuous series).

20. Why do we need to present data graphically ?

21. What are the different ways of presenting data graphically ? Suggest the best manner of graphically presenting data on:

(a) Trends in India's population over the last thirty years.

(b) Relative size of various states of India.

(c) Performance of the economy during this year relative to the last year as indicated by industrial production, price level and exports.

(Hint: (a) Time-series graph-years on x-axis and population on y-axis-make use of false base line (b) Better present it by a diagram-pie-diagram. (c) Time-series graph using three different variables to 100. Changes in the current year should be made accordingly.)

22. A survey showed that the average daily expenditure (in rupees) of 30 households in a city were:

11, 12, 14, 16, 16, 17, 18, 18, 20, 20, 20, 21, 21, 22, 22, 23, 23, 24, 25, 25, 26, 27, 28, 28, 31, 32, 32, 33, 36, 38,

(a) Form a frequency distribution using class intervals

10-14, 15-19, 20-24, 25-29, 30-34, 35-39

(b) What per cent of the households spend more than Rs.29/- each day ?

(c) Draw a frequency histogram for the above data.

(Hint: (a) With tally bar, prepare frequency table. (b) out of 30, 24 spent upto Rs.29/- and 6 spend more than 29/-. This means 20% (6 as % of 30) spent more than Rs.29/-. (c) For preparing histogram convert inclusive series into exclusive series).

23. For certain households, we have the following data on monthly income and expenditure:-

Income (Rs)	Expenditure (Rs)
100	95
150	135
250	210
400	300
600	420
1000	600

For each of the six households, calculate expenditure as a percentage of income. Plot these percentages on graph paper against income (choosing an appropriate scale).

What inference would you draw from the graph thus obtained ?

(Hint:

Income	Expenditure	Expenditure as % of Income
100	95	$\frac{100}{100} \times 95 = 95$
150	135	$\frac{135}{150} \times 100 = 90$
250	210	$\frac{210}{250} \times 100 = 84$
400	300	$\frac{300}{400} \times 100 = 75$
600	420	$\frac{420}{600} \times 100 = 75$
1000	600	$\frac{600}{1000} \times 100 = 60$

It is not possible to graph the required information.

It can be possible only with three dimensional diagram which is not called for here. The pattern of distribution of income indicates with increase in income the expenditure in absolute terms increases but in terms of percentages it declines).

24. What are different types of graphs ? Draw a histogram from a given data relating to the monthly pocket money allowance of the students of class XI of a school:-

0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
5	10	15	20	25	15	10	5

25. Distinguish between frequency polygon and Histogram. Present the data given in the table below in a Histogram:-

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	4	10	16	22	20	18	8	2

26. What points should be taken into account in the construction of graphs ?

27. What is false base line ? Under what conditions would its use be desireable?

28. Explain the general rule to construct a graph.

29. What is a smoothed frequency curve ? Discuss briefly various types of frequency curves.

30. The following data relates to the per capita net national product at 1970 -71 prices:

1974-75	1975-76	1976-77	1977-78	1978-79	1979-80	1980-81
618	661	653	695	717	661	696

Plot the above data on a graph paper using false base line.

UNIT IV

1. **Measures of Central Tendency - Averages**
2. **Measures of Central Tendency - Median**
3. **Measures of Central Tendency - Mode**
4. **Partition Values**
5. **Measures of Variability**
6. **Measures of Correlation (Two variables and Ungrouped Data)**
7. **Index Numbers**

1

Measures of Central Tendency-Averages

In every day life,we talk of average production, average expenditure or consumption, mean income, mean age, average weight and height, and so on. We also talk of per capita income of India, average size of land holdings belonging to an Indian farmer, average productivity of labour, average cost of production of various articles, and average life of an Indian etc. The concept of average is an important tool in the hands of statisticians for the analysis of statistical data. The averages are also called the measures of central tendency or location.

"An average is a significant single expression representative of the whole distribution". A measure of Central tendency will represent the whole of the distribution.

There are various types of measure of central tendency but for our purpose, the following are most important:-

1 Arithmetic Mean.

2 Median.

3 Quartiles, Deciles and Percentiles.

4 Mode.

For an ideal measure of central tendency, following requisites are necessary:-

i) It should be rigidly defined.

ii) It should be based on all observations.

iii) It should be representative of the whole distribution.

iv) It should be easily calculable.

v) It should not be influenced by the extreme values.

vi) It should be capable of mathematical calculations.

vii) It should not be affected by sampling fluctuations.

viii) It should be capable of easy interpretation.

ARITHMETIC MEAN

Arithmetic mean is generally abbreviated as A.M. It is defined as "*the quantity which is obtained by dividing the sum total of the various values of the variable by the number of such values*". It is denoted by $\overline{X}$. To calculate A.M. we use the following formula:-

$$\overline{X} = (\text{A.M.}) = \frac{X_1 + X_2 + X_3 + \ldots X_N}{N} = \frac{\Sigma X}{N}$$

Let us take an example to calculate Arithmetic Mean.

Illustration No. 1

The following are the monthly salaries in rupees of the employees of Delhi State Cooperative Stores, New Delhi. Calculate the arithmetic mean of their salaries:

Rs. 160, 187, 229, 245, 250, 300, 325, 350, 375, & 800.

Steps in Calculation

(a) Let variable X denote the salary of employee . To find A.M., we have to calculate ΣX (total salaries) and then divide the sum by 10 (i.e. the number of employees).

(b) X

160

187

229

245

250

300

325

350

375

800

3221

$$\overline{X} = \frac{\Sigma X}{N} = \frac{3221}{10}$$

$$= 332.10$$

Thus, $\overline{X}$ = 322.10 Ans.

Short-cut Method

Where the number of items i.e. N is very large, we can make use of the short-cut method. This method is explained in illustration no. 2.

Illustration No. 2

Following are the monthly expenditure of 12 families in Delhi, Calculate the arithmetic mean with short-cut and direct methods :

73, 77, 97, 100, 125, 128, 132, 133, 135, 140, 141, 168.

Short Cut Method

Steps in Calculation

a) Denote the items as X.

b) Assume any item of the series as mean and denote it as A.

c) Deduct this assumed mean from each item and denote it as dx.

d) Add all the dx and denote it as Σdx.

e) Apply the following formula:

$$\overline{X} = A \pm \frac{\Sigma dx}{N}$$

X	(X - A) dx
	(Suppose A = 128)
73	– 55
77	– 51
97	– 31
100	– 28
125	– 03
128	0
132	4
133	5
135	7
140	12
141	13
168	40
	$\Sigma dx = -87$

$$\overline{X} = A \pm \frac{\Sigma dx}{N}$$

$$= 128 - \frac{87}{12}$$

$$= 128 - 7.25$$

$$= 120.75$$

Thus $\overline{X}$ = 120.75 Ans.

Let us solve the illustration no. 2 by Direct Method.

X
73
77
97
100
125
128
132
133
135
140
141
168
1449

Applying formula:

$$\overline{X} = \frac{\Sigma X}{N}$$

$$= \frac{1449}{12}$$

$$= 120.75$$

Thus $\overline{X}$ = 120.75

You must have seen that the Arithmetic Mean is the same in both the cases i.e. the Direct and the Short-cut methods. It may look that the procedure to calculate arithmetic mean by short-cut method is lengthy. But in fact it is not. We shall see later that it is the most useful method.

Calculation of Mean in a Discrete Series

In discrete series, the values of the variables are multiplied by the respective frequencies, the products added together and the total divided by the total of the items (frequencies) i.e. Σf. We shall use the following formula for calculating arithmetic mean in a discrete series:

$$\overline{X} = \frac{\Sigma fx}{N} \quad \text{(Direct Method)}$$

$$\text{or } \overline{X} = A \pm \frac{\Sigma fdx}{N}$$

(Short cut Method)

$$\text{or } \overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f}$$

Now, let us take an illustration to show the calculation of arithmetic mean of discrete series by direct method and short-cut method

Illustration No. 3

Calculate the arithmetic mean from the following frequency distribution in a test in economics wherein 100 students participated.

Marks	5	10	15	20	25	30	35	40
No. of students (frequency)	12	15	13	18	11	09	08	14

Solution

Calculation of A.M. (Direct Method)

x	f	fx
5	12	60
10	15	150
15	13	195
20	18	360
25	11	275
30	09	270
35	08	280
40	14	560
	$\Sigma f = 100$	$\Sigma fx = 2150$

Steps in calculation

1. Denote the items by x and number of students by f.
2. Multiply respective x by f and denote the product as fx
3. Add column fx and denote the total as Σfx.
4. Apply formula-

$$\overline{X} = \frac{\Sigma fx}{\Sigma f}$$

$$\overline{X} = \frac{2150}{100} = 21.50 \text{ Ans.}$$

Short-cut method

We can calculate the arithmetic mean of the above illustration by short-cut method also. Let assumed mean A = 25

x	f	dx = (X-A)	fdx
5	12	– 20	– 240
10	15	– 15	– 225
15	13	– 10	– 130
20	18	– 05	– 90
25	11	0	0
30	09	+ 05	+ 45
35	08	+ 10	+ 80
40	14	+ 15	+ 210
	Σf = 100	Σfdx =	– 350

Steps in calculation

1. Denote the marks as x and number of students by f.
2. Assume any marks as assumed mean and denote it by A.
3. Find out dx by subtracting A from x.
4. Multiply respective f with dx.
5. Sum up the column of fdx and column at f.

6. Apply formula-

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f}$$

$$\overline{X} = 25 - \frac{350}{100}$$

$$= 25 - 3.50$$

$$= 21.50$$

STEP DEVIATION METHOD

Illustration No. 4

Steps in calculation

1. Denote the items as x and frequency as f.

2. Assume any items of x series as assumed average and denote it as A.

3. Deduct A from x and denote it as dx.

4. Divide the deviations dx taken from A by a common factor i, to get step deviation and denote it as dx .

5. Multiply respective step deviation dx′ and frequency f to get the product fdx′.

6. Sum up column no. 2 to get Σf and column no. 5 to get Σfdx′.

7. Apply the formula-

$$\overline{X} = A \pm \frac{\Sigma fdx'}{\Sigma f} \times i$$

A = 25 i = 5

x	f	dx	dx′	fdx′
1	2	3	4	5
5	12	– 20	– 4	– 48
10	15	– 15	– 3	– 45
15	13	– 10	– 2	– 26
20	18	–5	–1	–18
25	11	0	0	0
30	09	+ 5	+1	+09
35	08	+10	+2	+16
40	14	+15	+3	+42
	Σf = 100			Σfdx = – 70

Solution

$$\overline{X} = A \pm \frac{\Sigma fdx'}{\Sigma f} \times i$$

$$= 25 - \frac{70}{100} \times 5$$

$$= 25 - 3.5 = 21.50 \text{ Ans.}$$

Computation of A.M of a Continuous Frequency Distribution

In order to calculate A.M. for a frequency distribution of a continuous variable, we convert the class interval (C.I.) into discrete series by taking the mid-values of the class interval. Mid-values are obtained by adding the lower limit and upper limit of the C.I. and then dividing it by 2. Mid-value of C.I. is-

$$= \frac{\text{Lower limit of C.I.} + \text{Upper limit of C.I.}}{2}$$

Mid-value of the class interval is obtained because it is considered as the representative value of the class interval.

Illustration No. 5

Let us calculate the arithmetic mean from the following series:

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	5	12	18	22	6	4	3

DIRECT METHOD

Steps in calculation

1. Write the class intervals as in column no. 1

2. Obtain the mid-values of the class interval and denote the mid-value as x.

3. Multiply the mid-value with frequency and get fx.

4. Total up column no.3 to get Σf and total up column no.4 to obtain Σfx.

5. Apply the formula-

$$\overline{X} = \frac{\Sigma fx}{\Sigma f}$$

Solution

(Illustration No.5)

C.I (1)	M.V.(x) (2)	f (3)	fx (4) (2)x(3)
0-10	5	5	25
10-20	15	12	180
20-30	25	18	450
30-40	35	22	770
40-50	45	6	270
50-60	55	4	220
60-70	65	3	195
		$\Sigma f = 70$	$\Sigma fx = 2110$

$$\overline{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{2110}{70}$$

$$= 30.14 \text{ Ans.}$$

SHORT CUT METHOD

Steps in calculation

1. Write the class interval as in column no.1.
2. Take mid-values of the class intervals and denote it by x.
3. Assume any value from mid-value as A.
4. Take deviation from the assumed mean and denote it as dx as in column no.4.
5. Multiply the deviations with frequency and denote it as fdx.
6. Total up column no.3 to get Σf and column no.5 to get Σfdx.
7. Apply the formula-

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f}$$

Solution (A = 35)

C.I. (1)	M.V. x (2)	f (3)	dx (4)	fdx (5)
0-10	5	5	– 30	–150
10-20	15	12	–20	–240
20-30	25	18	–10	–180
30-40	35	22	0	0
40-50	45	6	+10	+60
50-60	55	4	+20	+ 80
60-70	65	3	+30	+ 90
		Σf = 70		Σfdx = –340

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f}$$

$$= 35 - \frac{340}{70}$$

$$= 35 - 4.86$$

$$= 30.14 \text{ Ans.}$$

STEP DEVIATION METHOD

Steps in Calculation

1. Take mid-values of Class Intervals and denote them by x.

2. Assume any item from mid-values as assumed mean and denote it by ' A'.

3. Take deviations from A and denote it as dx.

4. Divide deviation by ' i ', common factor to get step deviations and denote it by dx′.

5. Multiply the dx′ by respective frequencies and write it as fdx′.

6. Sum up column no.3 to get Σf and column no.6 to get Σfdx′.

7. Apply the formula-

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f} \times i$$

A = 35 i = 10

C.I. (1)	M.V. (2)	f (3)	dx (4)	dx′ (5)	fdx′ (6)
0-10	5	5	–30	– 3	– 15
10-20	15	12	– 20	– 2	– 24
20-30	25	18	– 10	– 1	– 18
30-40	35	22	0	0	0
40-50	45	6	+ 10	+ 1	+ 6
50-60	55	4	+ 20	+ 2	+ 8
60-70	65	3	+ 30	+ 3	+ 9
		Σf = 70			Σfdx′ = –34

$$\overline{X} = A \pm \frac{\Sigma\, fdx'}{\Sigma f} \times i$$

$$= 35 - \frac{34}{70} \times 10$$

$$= 35 - \frac{34}{7}$$

$$= 35 - 4.86$$

$$= 30.14 \text{ Ans.}$$

Merits and Demerits of Arithmetic Mean

MERITS

1. It is rigidly defined, thus, leaves no scope for different interpretations.
2. It is easy to calculate and understand.
3. It is based on all the items of the data and hence it is representative.
4. It is least affected by sampling fluctuations.
5. It is a more reliable basis for comparison.
6. It can be calculated without arranging the data in any form.

DEMERITS

1. It is very much affected by the existence of extreme values of the variables.
2. It is difficult to calculate in an open end classes.
3. It cannot be calculated in the absence of any of the items.
4. It can be a value which does not exist in the series at all.

CHARLIER' S CHECK FOR ACCURACY

The methods of calculating arithmetic mean by short-cut or step deviation methods are quite lengthy and there are possibilities of committing mistakes. Prof. Charlier has given a formula to check the accuracy of calculating arithmetic mean. The formulae are-

1. $\Sigma f(d+1) = +\Sigma f$(for short-cut Mehtod)
2. $\Sigma f(d'+1) = +\Sigma f$(for step deviation method)

Weighted Arithmetic Mean

A simple arithmetic mean is calculated when all the items have equal importance or weight, but when values of items differ in importance, we attach with them a certain weight according to their importance. In these cases, we make use of the weighted arithmetic mean. Suppose, Km. Sneh Lata has obtained 60, 75, 63, 60 and 50 marks in English, Commerce, Economics, Mathematics and Sanskrit. The Central Board of Secondary Education has allowed weights to these subjects as 1, 2, 2, 3, 1. We want to know her mean marks.

Solution

Subject (1)	Marks x (2)	Weight W (3)	Products Wx (4)
English	60	1	60
Commerce	75	2	150
Economics	63	2	126
Mathematics	60	3	180
Sanskrit	50	1	50
		$W = \overline{9}$	$Wx = \overline{566}$

$$\text{Weighted Mean} = \frac{\Sigma Wx}{\Sigma W}$$

$$= \frac{566}{9} = 62.89$$

$$= 62.89 \text{ Ans.}$$

Calculation of Mean when Mid-values are given

Illustration No.6

The distance of 100 picnic spots around Delhi is measured from Cannaught Place in kilometres as follows:

Mid-Values:	5	15	25	35	45
No. of picnic spots	12	18	34	20	16

Calculate the average distance.

Solution

The average distance can be calculated with the help of the arithmetic mean. We are already given the mid-values. We can calculate the class-intervals. The first class-interval will be-

$$(\text{mid-value} \pm \frac{1}{2} \text{ x i})$$

Where i stands for size of the class-interval. In our illustration, it will read as follows:

$$5 \pm 1/2(10) = 5 \pm 5 = 5 - 5 = 0$$
$$5 + 5 = 10$$
$$\text{C.I.} = 0\text{–}10$$

Similarly, other class-intervals will be:

10-20, 20-30, 30-40, 40-50 etc.

However, we are interested only in mid-values which are already given to us. We can calculate the mean now in the following way:

$A = 25, i = 10$

C.I	f	MV	dx	dx′	fdx′	
0-10	12	5	− 20	− 2	− 24	
10-20	18	15	−10	−1	−18	
20-30	34	25	0	0	0	− 42
30-40	20	35	+ 10	+ 1	+ 20	
40-50	16	45	+ 20	+ 2	+ 32	+ 52
	Σf = 100			Σfdx′ =	+ 10	

$$\text{Mean} = A \pm \frac{\Sigma fdx'}{\Sigma f} \text{ x i}$$
$$= 25 + \frac{10}{100} \text{ x } 10$$
$$= 25 + 1 = 26.0 \text{ Ans.}$$

Calculation of Mean for a Cumulative Frequency Distribution

Illustration No.7

Following is the distribution of marks obtained by 50 students in a test in Economics.

Marks (more than)	No.of students	Marks (more than)	No.of students
0	50	30	20
10	46	40	10
20	40	50	3

Calculate the arithmetic mean.

Solution

We are given the cumulative frequency series. All the 50 students have secured more than 0 marks: but 46 students have secured more than 10 marks and so on. It means that 4 students have secured marks between 0-10. We can similarly rearrange the distribution as follows:

Marks	No. of Students
0-10	4
10-20	6
20-30	20
30-40	10
40-50	7
50-60	3

From the above rearranged distribution we can calculate the mean in the following way:

$A = 25 \quad i = 10$

C.I.	f	MV	dx	dx'	fdx'	
0-10	4	5	− 20	− 2	− 8	
10-20	6	15	−10	−1	− 6	
20-30	20	25	0	0	0	
30-40	10	35	+ 10	+ 1	+ 10	−14
40-50	7	45	+ 20	+ 2	+ 14	
50-60	3	55	+ 30	+ 3	+ 9	+ 33
	Σf = 50				Σfdx' = 19	

$$\text{Mean} = A \pm \frac{\Sigma fdx'}{\Sigma f} \times i$$

$$= 25 + \frac{19}{50} \times 10$$

$$= 25 + \frac{19}{5}$$

$$= 25 + 3.8 = 28.8 \text{ Ans.}$$

Calculation of Mean for Open-End Classes

Illustration No.8

The following table gives the distribution of monthly income of 60 families in J.J.Colony, Delhi. Calculate the arithmetic mean.

Monthly income (Rs)					
below	75	75-150	150-225	225-300	300-375
No. of families	6	17	20	6	5
Monthly income (Rs): below	375-450	450 and over			
No. of families	4	2			

Solution

This is an illustration of an open-end class interval series. The lower limit of the first class-interval and the upper limit of the last class-interval is to be defined. These limits will be defined by their sizes. The magnitude of all other class-intervals is 75. Therefore,first class will be 0-75, and the last class will be 450-450 + 75 or 450-525. The arithmetic mean can be calculated in the following way:

Calculation of Mean for Open-End-Series

A = 262.5 i = 75

C.I.	f	MV	dx	dx'	fdx'	
0 - 75	6	37.5	−225	− 3	−18	
75 - 150	17	112.5	− 150	− 2	− 34	
150 - 22	20	187.5	−75	−1	−20	−72
225 -300	6	262.5	0	0	0	
300 -375	5	337.5	+ 75	+ 1	+ 5	
375 -450	4	412.5	+ 150	+ 2	+ 8	
450 -525	2	487.5	+ 225	+ 3	+ 6	+ 19
	Σf = 60				Σfdx' = −53	

$$\text{Mean} = A \pm \frac{\Sigma\, fdx'}{\Sigma f} \text{ x i}$$

$$262.5 - \frac{53}{4} \text{ x } 75 \quad = 262\,.5 - 66\,.25$$

$= 196\,.25$ Ans.

Calculation of Mean -Corrected Mean

Illustration No.9

The average weight of a group of 25 boys was calculated to be 78 .4 lbs. It was later discovered that one weight was misread as 69 lbs.

instead of 96 lbs. Calculate the correct mean.

Since $\overline{X} = \frac{\Sigma x}{N}$

Given $\overline{X} = 78.4$

$N = 25$

So $\Sigma x = \overline{X}N$

$= 78.4 \times 25 = 1960$

Subtracting the incorrect item and adding the correct item in x.

The correct $\overline{X} = 1960 - 69 + 96$
$= 1987$

The correct $\overline{X} = \frac{1987}{25}$

$= 79.48$

The above result can be obtained in another way also as follows:

$$\overline{X}_{\text{correct}} = \overline{X}_{\text{incorrect}} - \frac{\text{Incorrect item} - \text{Correct item}}{\text{No. of items}}$$

$$= 78.4 - \frac{69 - 96}{25} = 78.4 - \left(- \frac{27}{25}\right)$$

$$= 78.4 + \frac{27}{25}$$

$$= 78.4 + 1.08 = 79.48 \text{ Ans.}$$

Calculation of Combined Mean

When two or more distributions are given with their number of items and arithmetic means, we can calculate the combined mean as follows:

$$\overline{\overline{X}} \text{ (read } \overline{\overline{X}} \text{ double bar)} = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2}{N_1 + N_2}$$

Where $\overline{\overline{X}}$ = Combined Arithmetic Mean.

$\overline{X}_1$ = Arithmetic Mean of first distribution.

N_1 = No. of items in the first distribution.

$\overline{X}_2$ = Arithmetic mean of second distribution.

N_2 = No. of items in the second distribution.

Similiarly, when more than two distributions are given -

$$\overline{\overline{X}} = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2 + \dots\dots \overline{X}_n N_n}{N_1 + N_2 + \dots\dots N_n}$$

Illustration No.10

The Mean age of 40 students is 16 years and the mean age of another group of 60 students is 20 years. Find out the mean age of all the 100 students combined together.

Denoting the first group by 1 and second group by 2, we get-

$\overline{X}_1 = 16, N_1 = 40$

$\overline{X}_2 = 20, N_2 = 60$

$$\overline{X}_3 = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2}{N_2 + N_2}$$

$$= \frac{16 \times 40 + 20 \times 60}{40 + 60}$$

$$= \frac{640 + 1200}{100} = \frac{1840}{100} = 18.4$$

The combined mean age of 100 students is 18 .4

Calculation of Mean -Case of concealed frequency

Illustration No.11

The following table gives the distribution of 100 accidents during seven days of the week of a given month. During the particular month, there were 5 Mondays, Tuesdays and Wednesdays and only four each of the other days. Calculate the number of accidents per day.

Days	No.of Accidents	Days	No. of Accidents
Sunday	26	Thursday	8
Monday	16	Friday	10
Tuesday	12	Saturday	18
Wednesday	10		

Solution

Days	No. of Accidents	No. of Days in the month	xf
Sunday	26	4	104
Monday	16	5	80
Tuesday	12	5	60
Wednesday	10	5	50
Thursday	8	4	32
Friday	10	4	40
Saturday	18	4	72
		$\Sigma f = 31$	$\Sigma xf = 438$

$$\overline{X} = \frac{\Sigma xf}{\Sigma f} = \frac{438}{31}$$

$$= 14.1290$$

The average number of accidents per day is 14.1290

QUESTIONS

1. Define the mean of a frequency distribution. Calculate the mean of the following distribution of monthly wages of workers in a factory:-

Monthly wages (Rs)	No. of workers
100-120	10
120-140	20
140-160	30
160-180	15
180-200	5
	80

2. What is meant by Arithmetic Mean? State the formula to find it out. Explain its merits and demerits.

3. Define Mean. Why is Arithmetic Mean the most commonly used measure of central tendency? Compute A.M. from the following data.

Scores	Frequency
0-10	10
10-20	20
20-30	18
30-40	32
40-50	21

4. Calculate the Mean from the following data:

Marks	No. of students
0-10	4
10-20	6
20-30	8
30-40	10
40-50	7
50-60	6
60-70	5
70-80	3
80-90	1

5. Calculate arithmetic mean from the following data:-

Class	Frequency
20-25	10
25-30	12
30-35	8
35-40	20
40-45	11
45-50	4
50-55	5

6. Find the Arithmetic Mean from the following:

C.I.	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	3	7	16	12	9	5	3

7. Calculate the Mean-

Item	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	4	8	8	16	12	6	4

8. Compute the Mean

C.I.	0-20	20-40	40-60	60-80	80-100	100-120
Frequency	5	50	84	32	10	6

9. Calculate the Mean-

Age	16-19	20-29	30-39	40-49	50-59	60-69
Frequency	15	46	49	32	28	14

(Hint: Adjust class intervals as 15.5-19.5 and so on)

10. Calculate Mean from-

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	8	13	16	8	5

11. Find out Arithmetic Mean:-

C.I.	10-20	20-40	40-70	70-120	120-200
Frequency	4	10	26	8	2

(Hint: Use direct method after finding mid-values of class intervals)

12. Obtain the Mean of the population of 104 villages in a tehsil/taluk:-

Population	No. of villages
0 - 200	10
200 - 400	28
400-1000	42
1000-2000	18
2000-5000	6

(Hint: Find M.V.of C.I. of population. Assume mean, find deviation and proceed).

13. Calculate arithmetic mean:-

C.I.	5-15	15-25	25-35	35-45	45-55
Frequency	8	12	15	2	6

14. Find out the mean-

Item	Frequency
0-5	2
5-10	5
10-15	7
15-20	13
20-25	21
25-30	16
30-35	8
35-40	3

15. Given below are the values of a variable and its frequency. Calculate A.M. by direct and short-cut methods.

Size	No. of households
2	2
3	3
4	9
5	21
6	11
7	5
	51

16. Calculate mean form the following data on the monthly income (in rupees) of five households-852, 635,792, 836, 750

17. Find out the mean:

Marks obtained	No. of students
70-80	7
60-70	11
50-60	22
40-50	0
30-40	15
20-30	5

(Hint: Arrange C.I. as 20-30, 30-40 and so on).

18. Find out mean:-

Mid-points	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

(Hint: Find C.I. as 110-120 and so on)

19. Write a short note on weighted mean. Give its formula.

20. Find out the mean from the following:-

Rupees	No. of Receivers
Less than 10	15
Less than 20	35
Less than 30	60
Less than 40	84
Less than 50	96
Less than 60	127
Less than 70	190
Less than 80	250

(Hint: C.I. will be 0-10, 10-20 Frequency will be 15, 20 and so on)

21. Calculate the mean from the following series:-

C.I.	Frequency
5-15	8
15-25	12
25-35	15
35-45	9
45-55	6

22. Find the mean from the following

Less than-	7	14	21	28	35	42	49	56
Frequency	26	57	92	134	216	287	341	360

[Hint: arrange with suitable class intervals as 0 -7, 7 -14 and so on]

2

Measures Of Central Tendency - Median

Median is also a measure of central tendency. Arithmetic mean is affected by the values of the extreme items. In order to avoid the effect of the extreme values, *we take a middle position in the series.* This position is called Median. According to *Prof. H. Secrist "Median of a series is the value of that item, actual or estimated when a series is arranged in order of magnitude which divides the distribution into two parts."* Thus, median is that value of the series which divides the series into two equal parts, one part comprising of values less than and the other greater than it.

(a) Calculation of Median in Individual Series

To get median value in the individual series, we use the following formula:-

$M = \frac{N+1}{2}$ (N denotes number of the items)

Illustration No.1

Calculate Median from the following-

15, 20, 25, 28, 16, 18, 17, 9, 11.

Steps in calculation

1. First of all arrange the values of the variable in the ascending order.
2. Use the following formula:

$$M = \frac{N+1}{2}$$

Solution

x
9
11
15
16
17
18
20
25
28
N =9

$$M = \frac{N+1}{2} = \frac{9+1}{2}$$

= 5th item

= the value of the 5th item is 17

Thus, M = 17 Ans.

Note: In the above illustration, the number of items is odd(9).

In such a case, median can be easily calculated, but if the number of items is even i.e. 2, 4, or 6, median will be calculated in the following manner.

Illustration No.2

Given below is the age of some students. Find out the median of their age.

S.no.	1	2	3	4	5	6	7	8
Age(Yrs)	18	16	14	11	13	10	9	20

Steps in calculation

1. Arrange the values in ascending order as

1 2 3 4 5 6 7 8
9 10 11 13 14 16 18 20

2. As the number of items is even i.e. 8, the item will be

$$\frac{8+1}{2} = 4.5$$

3. We have to add the values of the 4th and 5th items and divide them by 2.
4. Thus, Median will be the value of

$$M = \frac{N+1}{2}$$

$$= \frac{8 \quad +1}{2} \quad = \quad \frac{9}{2} = 4.5 \text{th item}$$

$$= \frac{13 \quad +14}{2} = 13.5 \text{ years.}$$

$$M = 13.5 \text{ Ans.}$$

Calculation of Median from Discrete Series

In a discrete series, the values of the variable are given alongwith their frequencies. The calculation of median in such a series is illustrated below:-

Illustration No.3

Calculate Median from the following data-

Variable	5	10	15	20	25	30
Frequencies	18	1	1	9	13	8

Steps in calculation

1. Denote item as x and frequency as f.
2. Prepare the cumulative frequency (cf) as in column 3.
3. Find the median item by using-

$$M = \frac{N \ +1}{2} \text{th item}$$

4. To locate the value of median corresponding to $(N + 1/2)$th item, find out in which cumulative frequency column it falls, the corresponding value of the variable will be the median.

Solution

Variable (x)	Frequencies (f)	Cumulative frequencies (cf)
5	18	18
10	16	34
15	14	48
20	9	57
25	13	70
30	8	78

$$M = \text{Value of } \frac{N+1}{2} \text{ th item}$$

$$= \text{Value of } \frac{78+1}{2} \text{ th item}$$

$$= \text{Value of } 39.5\text{th item}$$

$$= 15 \text{ Ans.}$$

Calculation of Median in a Continuous Series

In a class-interval series, median item can be found as N/2. The median item so determined will be in a class interval which is called median class and the value of median is obtained with the help of the following formula:-

$$\text{Median} = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

In the above formula-

l_1 is the lower limit of the median group.

i- is the width of the class interval in which median lies.

N- is the total number of frequencies.

f- is the frequency of the median class.

C- cumulative frequency of the class preceeding the class in which median lies.

The above formula is applied when the series is in ascending order. In case the series is given in descending order, it is converted into ascending order and the above formula will be applied.

Illustration No.4

Calculate median from the following data:-

Age (in years)	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons	50	70	100	180	150	120	70	60

Steps of calculation

1. Calculate cf.
2. Determine the class by using N/2 in which the median lies.
3. Find the median by applying the formula.

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

Solution

C.I.	F	cf
20-25	50	50
25-30	70	120
30-35	100	220
35-40	180	400
40-45	150	550
45-50	120	670
50-55	70	740
55-60	60	800
	800	

Size of median item-

$$M = \frac{N}{2}$$

$$= \frac{800}{2}$$

$$= 400$$

Thus, 35-40 is median size class interval.

By applying formula-

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 35 + \frac{400 - 220}{180} \times 5$$

$$= 35 + \frac{180}{180} \times 5$$

$$= 35 + 5$$

$$= 40 \text{ years.}$$

Illustration No.5

Calculate Median from the following:-

C.I.	60-69	50-59	40-49	30-39	20-29	10-19
Frequencies	13	15	21	20	19	12

Steps of calculation

1. In the above question, the series is inclusive and is in descending order. So, we have to convert the series into exclusive and ascending order.

2. Repeat the steps as given in illustration no.4 above.

Solution

C.I.	f	cf
9.5-19.5	12	12
19.5-29.5	19	31
29.5-39.5	20	51
39.5-49.5	21	72
49.5-59.5	15	87
59.5-69.5	13	100

Size of median item-

$$M = \frac{N}{2}$$

$$= \frac{100}{2} = 50$$

Thus 29.5 – 39.5 is the median size class-interval.

By applying formula-

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 29.5 + \frac{\frac{100}{2} - 31}{20} \times 10$$

$$= 29.5 + \frac{50 - 31}{20} \times 10$$

$$= 29.5 + 9.5$$

$$= 39 \text{ Ans}$$

Illustration No.6

Calculation of Median when mid-values are given

Mid-values:	20	30	40	50	60	70	80
Frequency:	8	13	12	18	16	6	7

Steps of calculation

1. Convert the mid-values in class intervals.

2. When class intervals have been obtained, repeat the steps given in illustration no.4 above.

Solution

Mid value	C.I.	f	cf
20	15-25	8	8
30	25-35	13	21
40	35-45	12	33
50	45-55	18	51
60	55-65	16	67
70	65-75	6	73
80	75-85	7	80

Median size item-

$$M = \frac{N}{2}$$

$$= \frac{80}{2} = 40$$

Thus, Median lies in C.I. 45-55

Applying formula

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 45 + \frac{\frac{80}{2} - 33}{18} \times 10$$

$$= 45 + \frac{40 - 33}{18} \times 10$$

$$= 45 + \frac{70}{18}$$

$$= 45 + 3.89$$

$$= 48.89 \text{ Ans.}$$

Calculation of Median from unequal Class-Intervals

When the class intervals are unequal, it is needless to convert the C.I. into equal class intervals. The median can be calculated from the unequal class intervals. The only care that is to be taken is in respect of width of class interval'i'. The value of'i'should be the value of where median is located.

Illustration No.7

C.I.	0-5	5-10	10-20	20-30	30-50	50-70	70-100
Frequency	12	15	25	40	42	14	8

Solution

C.I	f	cf
0-5	12	12
5-10	15	27
10-20	25	52
20-30	40	92
30-50	42	134
50-70	14	148
70-100	8	156

Size of Median item

$$M = \frac{N}{2} = \frac{156}{2} = 78$$

So, Median lies in 20-30 C.I.

By applying formula-

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 20 + \frac{\frac{156}{2} - 52}{40} \times 10$$

$$= 20 + \frac{78-52}{40} \times 10$$

$$= 20 + \frac{26}{40} \times 10$$

$$= 20 + 6.5$$

$$M = 26.5 \text{ Ans.}$$

Calculation of Median from Cumulative frequencies

When cumulative frequencies less than or more than type are given, we convert them into regular class intervals and then follow the above steps.

Illustration No.8

Calculate median from the following cumulative distribution:-

Marks (less than):	15	30	45	60	75	90
No. of students:	18	35	62	81	95	100

Solution

C.I.	F	cf
0-15	18	18
15-30	17	35
30-45	27	62
45-60	19	81
60-75	14	95
75-90	5	100

Median size item-

$$M = \frac{N}{2} = \frac{100}{2} = 50$$

So, 30-45 is the C.I. of median size.

Applying formula-

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 30 + \frac{\frac{100}{2} - 35}{27} x\ 15$$

$$= 30 + \frac{15}{27} \quad x\ 15$$

$$= 30 + 8.33$$

$$= 38.33 \text{ Ans.}$$

Illustration No. 9 (Cumulative distribution-More than)

Weight in Kgs	70	65	60	55	50	45	40
No. of persons	8	19	42	63	85	98	100

Solution

As the cumulative distribution is in descending order, we have to convert the series into ascending order.

A

C.I.	f
70-75	8
65-70	11
60-65	23
55-60	21
50-55	22
45-50	13
40-45	2

B

C.I.	f	cf
40-45	2	2
45-50	13	15
50-55	22	37
55-60	21	58
60-65	23	81
65-70	11	92
70-75	8	100

Median size item-

$$M = \frac{N}{2} = \frac{100}{2} = 50$$

So, Median size C.I. is 55-60

Applying formula-

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 55 + \frac{\frac{100}{2} - 37}{21} \times 5$$

$$= 55 + \frac{13}{21} \times 5 \qquad = 55 + \frac{65}{21} = 55 + 3.095$$

M = 58.095 Kg. Ans.

Calculation of Median with the help of Graph

Median can also be calculated with the help of Cumulative Frequency Curve or *Ogive*. The following steps will be taken to calculate median with the help of OGIVE.

1. Draw a less than OGIVE or greater than OGIVE.

2. Compute N / 2 and locate this point on y axis.

(representing cumulative frequency).

3. Draw a line parallet to x axis (where the values of the variable are represented) at a distance of N / 2 i.e. draw a line y = N /2 to meet the OGIVE.

4. From the point of intersection, drop a perpendicular to the x axis. The intersection gives the value of median.

Illustration No.10

Draw the cumulative frequency curve from the following data and find out the median.

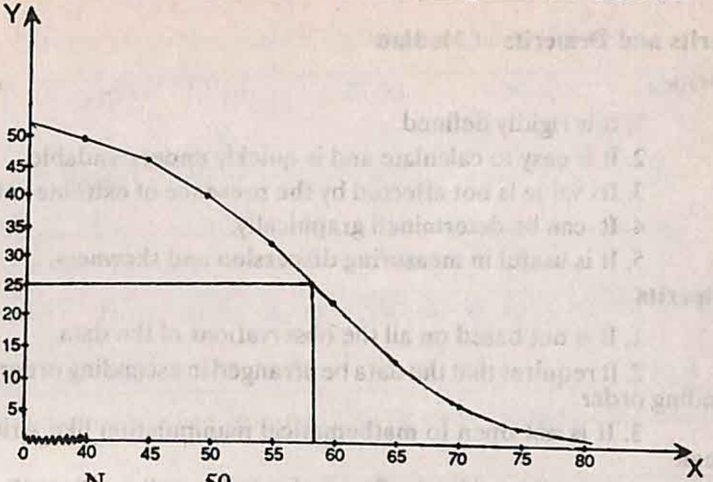
C.I	40-45	45-50	50-55	55-60	60-65	65-70
Frequency	4	6	8	10	7	6
C.I.	70-75	75-80	80-85			
Frequency	5	3	1			

Solution

In order to calculate median by cumulative frequency curve, we have to convert the series in cumulative frequency of 'more than' and 'less than' series.

C.I	Less than	cf	More than	cf
40-45	45	4	40	50
45-50	50	10	45	46
50-55	55	18	50	40
55-60	60	28	55	32
60-65	65	35	60	22
65-70	70	41	65	15
70-75	75	46	70	9
75-80	80	49	75	4
80-85	85	50	80	1

MORE THAN CUMULATIVE FREQUENCY CURVE



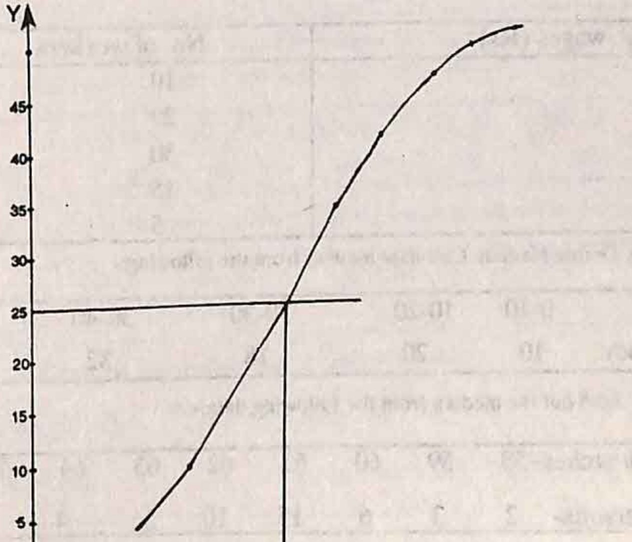
$$M = \frac{N}{2} = \frac{50}{2} = 25$$

$$M = l_1 \frac{\frac{N}{2} - C}{f} \times i$$

$$= 55 + \frac{\frac{50}{2} - 18}{10} \times 5$$

$$= 55 + \frac{25 - 18}{10} \times 5 \qquad = 55 + \frac{7}{2} \quad = 55 + 3.5 \quad = 58.5$$

LESS THAN CUMULATIVE FREQUENCY CURVE



Merits and Demerits of Median

Merits

1. It is rigidly defined.
2. It is easy to calculate and is quickly understandable.
3. Its value is not affected by the presence of extreme values.
4. It can be determined graphically.
5. It is useful in measuring dispersion and skewness.

Demerits

1. It is not based on all the observations of the data.
2. It requires that the data be arranged in ascending order or descending order.
3. It is not open to mathematical manipulation like Arithmetic Mean.
4. It is affected by the fluctuations of sampling. Thus, if class intervals are not uniform,the value of median becomes inappropriate.

Median is mostly used (i) to calculate the intelligence of the students, and (ii) to calculate the relative standing of an individual in a certain class or a society.

QUESTIONS

1. What is meant by Median? State the formula to find it out.Also explain its merits and demerits.

2. Define the median of a frequency distribution. Calculate the median for the distribution of wages given below:-

Monthly wages (Rs)	No. of workers
100-120	10
120-140	20
140-160	30
160-180	15
180-200	5

3. Define Median. Calculate median from the following:-

Scores	0-10	10-20	20-30	30-40	40-50
Frequency	10	20	18	32	21

4. Find out the median from the following data:-

Height in inches-	58	59	60	61	62	63	64	65	66
No. of persons-	2	3	6	15	10	5	4	3	1

5. Calculate the Arithmetic Mean, Median from the following data:-

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	4	6	8	10	7
C.I.	50-60	60-70	70-80	80-90	
Frequency	6	5	3	1	

6. Calculate the Median by using the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No.of students:	3	10	17	7	6	4	2	1

7. Draw 'Less than' and 'More than' OGIVE curves and find out the value of the median-

0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
7	10	20	13	12	10	14	9

8. Calculate the Median from the following:-

Marks (below)	10	20	30	40	50	60	70	80
No.of students	15	35	60	84	96	127	198	250

9. The percentage of marks obtained by 68 students in an examination are given below. Compute the Median:-

Marks-	0-20	20-40	40-60	60-80	80-100
No. of students	0	5	22	25	16

10. Calculate the Median from an analysis for more efficiency in a factory indicating the distribution of ages of workers which was as follows:

Age	10-19	20-29	30-39	40-49	50-59	60-69
Frequency	15	46	49	32	28	14

11. Calculate the median of the following frequency distribution:

C.I.	10-20	20-40	40-70	70-100	120-200
Frequency	4	10	26	8	2

12. Calculate the median from the following data:

Weekly wages (Rs)	58	59	60	61	62	63	64	65	66
No. of workers	2	3	6	15	10	5	4	3	1

13. Calculate median:-

Size	2	3	4	5	6	7
No. of house holds	2	3	9	21	11	5

14. Find out Median from the following-

Mid-points:	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

15. Calculate the value of Median:

Age	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons	50	70	100	180	150	120	70	60

16. The following are the marks of 50 students in statistics.Find Median-

Marks	5-15	15-25	25-35	35-45	45-55
No. of students	8	12	15	9	6

17. Find out the median from the following:-

Less than Rs.	10	20	30	40	50	60	70	80
No. of Receivers.	15	35	60	84	96	127	190	250

18. The percentage marks obtained by 100 students in an examination are given below. Calculate the median:-

Marks	30-35	35-40	40-45	45-50	50-55	55-60	60-65
No. of students	14	6	18	23	18	8	3

19. From the data given below, calculate the value of median:-

Income-	100	200	250	280	300	350	400
No. of persons	16	24	66	30	20	3	3

20. Calculate the median age from the following data:-

55-60	50-55	45-50	40-45	35-40
7	13	15	20	30

21. Explain the relative merits and demerits of mean and median.
(Hint: Relative merits and demerits of mean and median are:-

Mean is easy to understand and simple to calculate.It is based on all observations. It is rigidly defined.However,it is affected by extreme value. Median, though not affected by extreme values, is not based on all observations. It is not rigidly defined).

3

Measures Of Central Tendency - Mode

Mode is another important measure of Central Tendency. *It is defined as the value of variable which occurs maximum number of times. It is most common value found in a series.*

It is the point of greatest density, i.e., the items of the series are mostly concentrated around the mode. For this reason, the mode is also called the most fashionable value of the distribution. *According to Prof. Zizek "the value occurring most frequently in a series (or group) of items and around which the items are distributed most densely is called the Mode".* This value is denoted by the *symbol "Z"*.

Calculation of Mode

Individual series:

In an individual series mode can be found out by determining the value which occurs maximum number of times.

Illustration No.1

Calculate the value of mode from the following data:-

41, 42, 45, 44, 45, 48, 50, 45, 47, 51, 56

Solution

By observation we notice that 45 has highest frequency of 3. Thus 45 is mode.

Illustration No.2

In the series 27, 28, 26, 35, 40, 37, 38, 28, 27, 25, 30, 28, 40.

We observe that 28 has occurred 3 times while 27 and 40 have occurred 2 times each. Thus, 28 is the mode in this series.

Discrete Series

In a discrete series also mode can be found by observation.

Value with the highest frequency will be called mode.

Illustration No.3

Estimate the value of mode

X	21	22	25	26	27	28	29	30
f	7	10	15	18	13	7	3	2

Solution

We observe that in this series 18 is the highest frequency corresponding to the value of 26. Hence mode is 26.

Illustration No.4

Find mode from the following:

X	10	20	30	40	45	50	55	60
f	28	73	115	205	110	88	37	13

Solution

We observe that maximum frequency is 205 relating to 40. Hence, mode is 40.

GROUPING TABLE

In some cases it is possible that the value having the highest frequency may not be the mode. This will specially be so when the frequencies in the immediate neighbourhood of highest frequency are quite low. If we determime the mode by observation or inspection we may commit mistake. For example,let us look at the following data:

Illustration No.5

Marks	1	2	3	4	5	6	7	8	9	10	11	12
No. of students	4	7	8	12	16	14	9	7	17	5	3	2

According to observation, marks 9 has the frequency of 17 so by observation we will say 9 is mode. But according to the definition of Zizek 9 is not the mode because though 9 is occurring maximum number of times yet we see that around 5 the items are most densely distributed and around 9 it is not so. To overcome this difficulty, we use another method and that is known as Grouping Method. This method is illustrated below.

1. In column I we write the original frequencies against the respective items.

2. In column II, we add the frequencies in two's beginning with first item.

3. In column III, we add the frequencies in two's beginning with second item.

4. In column IV, we add the frequencies in three's beginning with first item.

5. In column V, we add the frequencies in three's beginning with second item.

6. In column VI, we add the frequencies in three's beginning with third item.

7. Then we prepare an analysis table and note down the items having maximum frequencies in each column of the grouping table.

8. We add up the columns of analysis table, the item which occurs maximum number of times will be mode.

Group Table

Variable items	I	II	III	IV	V	VI
1	4					
2	7	11				
3	8		15	19		
4	12	20			27	
5	16		28			36
6	14	30		42		
7	9		23		39	
8	7	16				30
9	17		24	33		
10	5	22			29	
11	3		8			25
12	2	5		10		

Analysis Table

Column. No	Variable item having maximum frequency						
1							9
2			5	6			
3		4	5				
4		4	5	6			
5			5	6	7		
6	3	4	5				
	1	3	5	3	1	x	1

From the above analysis table, it is clear that "5" occurs maximum number of times. Thus 5 is mode.

Class Interval Series or Continuous Series

In the case of grouped frequency distribution we first find out the Modal Class by locating either by observation or with the help of grouping table. The class in which mode lies is called the modal class.

After finding out the modal clas we use the following formula to locate the mode:

$$Z = l_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

l_1 is the lower limit of the modal class.

f_m is the frequency of the modal class.

f_1 is the frequency of pre-modal group.

f_2 is the frequency of post modal group.

'i' is the width of the modal class.

Illustration No.6.

Calculate the value of Mode.

C.I.	0-10	10-20	20-30	30-40	40-50
f	4	6	5	10	20
C.I.	50-60	60-70	70-80	80-90	90-100
f	22	24	6	2	1

Solution

For finding the modal class we use the grouping table because the group 60-70 which has frequency of 24 is followed by a group (70-80) which has frequency only 6, thus we will make a mistake according to Zizek's definition if we select a class which is followed by less densely distribution.

Grouping table

Class Interval	I	II	III	IV	V	VI
0-10	4					
10-20	6	10				
20-30	5		11	15		
30-40	10	15			21	
40-50	20		30			35
50-60	22	42		52		
60-70	24		46		66	
70-80	6	30				52
80-90	2		8	32		
90-100	1	3			9	

Analysis Table

Col.No.					
I				60-70	
II		40-50	50-60		
III			50-60	60-70	
IV	30-40	40-50	50-60		
V		40-50	50-60	60-70	
VI			50-60	60-70	70-80
	1	3	5	4	1

Since 50-60 has got maximum frequency it is modal group.

By applying the formula:

$$Z = l_1 + \frac{fm - f_1}{2fm - f_1 - f_2} \times i$$

$$= 50 + \frac{22-20}{2 \times 22 - 20 - 24} \times 10$$

$$= 50 + \frac{2}{44 - 44} \times 10$$

$$= 50 + \frac{2}{0} \times 10$$

$$= 50 \times \alpha$$

$$= \alpha$$

Hence this formula is not valid for this problem. We shall use the second formula to calculate mode.

The formula is

$$l_1 + \frac{f_2}{f_1 + f_2} \times i$$

$$= 50 + \frac{24}{20 + 24} \times 10$$

$$= 50 + \frac{240}{44} \qquad = 50 + 5.45 = 55.45$$

Illustration No.7

Find out mode from the following frequency distribution.

Mid-value	1	2	3	4	5	6	7	8	9	10
Frequency	8	6	10	12	20	12	5	3	2	4

Solution

From the mid value prepare class intervals

Mid value	C.I.	f
1	.5-1.5	8
2	1.5-2.5	6
3	2.5-3.5	10
4	3.5-4.5	12
5	4.5-5.5	20
6	5.5-6.5	12
7	6.5-7.5	5
8	7.5-8.5	3
9	8.5-9.5	2
10	9.5-10.5	4

By grouping table find the modal group

C.I.	I	II	III	IV	V	VI	Tally Total
.5 -1.5	8						
1.5-2.5	6	14					
2.5-3.5	10		16	24			1
3.5-4.5	12	22			28		3
4.5-5.5	20		32			42	6
5.5-6.5	12	32		44			3
6.5-7.5	5		17		37		1
7.5-8.5	3	8	5			20	
8.5-9.5	2			10			
9.5-10.5	4	6			9		

Hence 4.5 to 5.5 is a modal group as it has the highest frequency of 6 .

By applying formula:

$$Z = l_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$= 4.5 + \frac{20-12}{2 \times 20-12-12} \times 1$$

$$= 4.5 + \frac{8}{16} \times 1$$

$$= 4.5 + .5 = 5$$

Illustration No. 8

Find out mode of the following distribution.

C.I.	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	3	7	16	12	9	5	3

Solution

By observation we can say 20-25 is the modal group. By applying formula:

$$Z = l_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$= 20 + \frac{16 - 7}{2 \times 16 - 7 - 12} \times 5$$

$$= 20 + \frac{9}{13} \times 5$$

$$20 + \frac{45}{13}$$

$$= 20 + 3.46$$

$$= 23.46$$

Illustration No.9

Compute mode from the following data:

C.I.	0-10	10-20	20-30	30-40	40-50
F	10	20	18	32	21

Solution

By observation we find that modal group is 30-40. By applying formula:

$$Z = l_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$= 30 + \frac{32 - 18}{2 \times 32 - 18 - 21} \times 10$$

$$= 30 + \frac{14}{25} \times 10$$

$$= 30 + \frac{28}{5} = 30 + 5.6$$

Ans. $Z = 35.6$

Location of Mode Graphically

Mode can be located graphically with the help of histogram . We first prepare the histogram taking C.I. on X axis and frequencies on Y axis. Thereafter, the top corner of the rectangle having highest frequency is joined with the right and left corners of the rectangles preceeding and succeeding the modal class (i.e having highest frequency). From the point of intersection of the lines draw a perpendicular on the X axis. The variable will be located on X axis.

Please note that this method is suitable only when the mode lies in the class interval having maximum frequency.

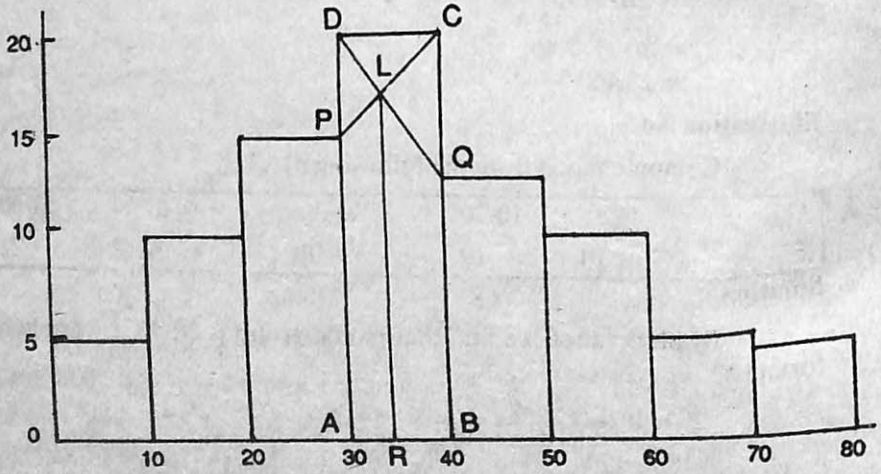
Illustration No. 10

Find out mode graphically and verify it by applying formula.

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	10	15	20	12	10	5	4

Solution

We prepare the histogram with the above C.I.'s and frequencies. On X axis are C.I.'s and on Y axis are frequencies.



In the above histogram rectangle ABCD represents the modal class . Point C & D are joined with the corners of the rectangles preceeding and succeeding respectively at points P & Q. The lines joining these ends intersect at L. From L perpendicular LR is drawn on X axis. OR represents Mode. The value is 34.

Let us verify the above from formula.

$$\text{Mode} = l_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$= 30 + \frac{20 - 15}{2 \times 20 - 15 - 12} \times 10$$

$$= 30 + \frac{50}{13}$$

$$30 + 3.84 = 33.84$$

Ans. 33.84 or say 34.

Mode can be calculated from Mean and Median

There is a relationship between Mode, Mean and Median.The relationship is as under.

Mode = 3 Median-2 Mean

Illustration No.11

In a symmetrical distribution the values of mean and median are 20.2 and 18.8 respectively. Find the value of mode:

Solution

Mean = 20.2 and Median = 18.8

Mode = 3 Median -2 Mean

$= 3 \times 18.8 - 2 \times 20.2$

$= 56.4 - 40.4$

$= 16.00$

Merits, Demerits, and Uses of Mode:

Merits:

1. It is easy to understand and can be located even by inspection.
2. It is not affected by the extreme items.
3. It can be found out by using graphs.
4. It does not need all the items.

Demerits

1. It is uncertain and ill-defined.
2. It is not based on all observations

3. It cannot be manipulated mathematically.
4. Different formulae give different results.

Use of Mode

It is widely used in production and trade. The manufacturer of shoes or readymade garments or the manufacturers of the cold drinks are very much interested to know the sizes or quality of their items liked by the public.

Relationship between Mean, Median and Mode

Mean, Median & Mode have their distinct role and they cannot be substituted for one another. All the three have some relationship with one another. However, mean is found by dividing the sum of the values by the number of items. Median is the value of the central item dividing the series into two extremes. Mode is the value most often found in the series. Their use depends upon the nature of the problem under investigation. If we want to know the average value,arithmetic mean is most suitable. The average helps to compare two different distributions relating to the same variable. We can compare the per capita income of India and U.S.A. We can know the level of economic development in two countries. Median helps us to understand those items which are not capable of measurement in definite units. It also helps us to avoid the effect of the extreme items.

Mode is useful when we are interested to know the maximum choice of the consumers. A readymade garment manufacturer and hosiery dealers are interested to know about the sizes and the colours etc. which are much in demand.

The three central tendency measures can be compared as under:

Comparison of three measures viz. Mean, Median and Mode

Basis of Comparison	Mean	Median	Mode
1. Definition	Rigid	Rigid	Not Rigid
2. Understanding	Easy	Easy	Easy
3. Calculation	Easy	Easy	Easy
4. Based on all observations	Yes	No	No
5. Capable of mathematical treatment	Yes	No	No
6. Affected by Extreme values	Yes	No	No
7. Computed graphically	No	Yes	Yes

QUESTIONS

1. What is meant by Mode? State the formula to find it out. Also explain its merits and demerits.

2. Calculate the mode from the following:

Income	100-200	200-300	300-400	400-500	500-600
Persons	15	18	30	20	17

3. If the values of mode and median are 40 and 44 respectively, find the value of mean.

4. Find out the mode from the following:

	Rs.	No. of Receivers	Rs.	No. of Receivers
Less than	10	15	Less than 50	96
	20	35	60	127
	30	60	70	190
	40	84	80	250

(Hint: Prepare C.I. as 0-10, 10-20..,70-80)

5. Find out mode from the following frequency distribution:

Mid value:	1	2	3	4	5	6	7	8	9	10
Frequency:	8	6	10	12	20	12	5	3	2	4

6. Find out the mode from the following distribution:

x	30-40	40-50	50-60	60-70	70-80	80-90	90-100
f	6	10	16	14	10	5	2

7. Find out the mode from the following

Marks obtained:	0-9	10-19	20-29	30-39	40-49
No. of candidates:	6	29	87	181	247
Marks obtained:	50-59	60-69	70-79	80-89	90-99
No. of candidates:	269	133	43	9	2

8. Draw a histogram from the following and indicate Mode therein

16-19	20-29	30-39	40-49	50-59	60-69
15	46	49	32	28	14

9. **Find the mode of the distribution from the following data:**

10-15	15-20	20-25	25-30	30-35	35-40	40-45
3	7	16	12	9	5	3

10. **Find the mode from the following data:**

Wages upto Rs.	No. of workers
10	12
20	30
30	65
40	107
50	157
60	202
70	222
80	230

(Hint: C.I. will be	0-10,	10-20,	20-30,	30-40,	40-50,	50-60.
Frequency	12	18	35	42	50	45
	60-70	70-80				
	20	8				

11. **Define Mode. Compute mode from the following data·**

Scores:	0-10	10-20	20-30	30-40	40-50
Frequency:	10	20	18	32	21

4

Partition Values

We have seen that median divides a series into two equal parts. But at times, we want to know the values, which divide the series into fours, tens and hundreds. *Quartiles, Deciles and Percentiles are such measures.*

QUARTILES

Quartiles divide a series into four equal parts. Thus for any distribution there are three quartiles denoted as Q_1, Q_2, Q_3.

Q_1 is also known as first or lower quartile which divides the distribution in such a way that one-fourth of total observations fall below it and three fourths are above it. It is also called first quartile.

Q_2 is known as median or middle quartile and it divides the distribution in such a way that one half of total observations fall below it and one-half is above it. It is called second quartile.

Q_3 is known as upper quartile. It divides the distribution in such a way that three fourths of total observations fall below it and one fourth above it. It is called third quartile.

Quartiles are located in the same way as median is located. In the individual and discrete series Q_1 is found by determining the size of $(N+1)/4$th item, and Q_3 is found by determining the size of $3(N+1)/4$th item.

For the class interval series we determine the Q_1 by first finding the value of N/4th item. The Q_3 is determined by 3N/4th item. After finding the size of the item of Q_1 and Q_3 we apply the following formulae to determine the value of Q_1 and Q_3.

$$Q_1 = l_1 + \frac{\frac{N}{4} - C}{f} \times i$$

$$Q_3 = l_1 + \frac{\frac{3N}{4} - C}{f} \times i$$

l_1 is the lower limit of the class interval in which the quartile lies.

C is the cf of the class interval preceeding the quartile class, 'i' is the width of the class interval containing quartile, f is the frequency of the class in which the quartile lies.

DECILES

It divides the distribution into ten equal parts. For any series there can be nine deciles-D_1 to D_9. The value of the fifth decile will be the median of the series.

The value of decile is the size of the $\frac{(N+1)\ J}{10}$ th item.

For grouped frequency distribution the deciles are calculated with the help of the following formula:

$$D_1 = l_1 + \frac{\frac{(N \times j)}{10} - C}{f} \times i$$

Here j = 1, 2, 3, 4, 5, 6, 7, 8, 9

l_1 is the lower limit of C.I. where the decile value lies. C is the cf of the CI precding the decile class.

f is the frequency of the C.I. where the value of decile lies.

'i' is the width of the C.I. where the value of decile lies.

PERCENTILES

Percentiles divides the series into hundred equal parts. For any series there can be 99 percentiles-P_1 to P_{99}. The value of the percentile is equal to the size of the $(N+1) \times k\ /\ 100$ th item. For grouped frequency distribution the percentiles are calculated with the help of the following formula:

$$P = l_1 + \frac{\frac{N \times k}{100} - C}{f} \times i$$

Where k = 1 to 99

l_1, i, f and c have meanings similar to the ones defined for deciles and quartiles.

INDIVIDUAL SERIES: QUARTILES

Illustration No.1

Calculate Q_1, Q_2, D_3, D_7, P_{35}, P_{80} from the following marks obtained by 19 students:

Serial No.	1	2	3	4	5	6	7	8	9	10
Marks:	18	20	25	17	9	11	23	37	38	42

Serial No.	11	12	13	14	15	16	17	18	19
Marks:	36	35	8	6	11	21	20	41	35

Solution:

First of all arrange the above marks in ascending order:

S.No.	Marks	S.No.	Marks	S.No.	Marks
1	6	8	20	14	35
2	8				
3	9	9	20	15	36
4	11	10	21	16	37
5	11	11	23	17	38
6	17	12	25	18	41
7	18	13	35	19	42

$$Q_1 = \frac{N+1}{4} \text{ th item's value.}$$

$$= \frac{19+1}{4} \text{ th item} = 5\text{th item. Value of 5th item} = 11$$

$$Q_3 = 3(N+1)/4\text{th items value.}$$

$$= \frac{3(19+1)}{4} = \frac{60}{4} = 15 \text{ th items value.}$$

Value of 15 th item $= 36$

$$D_3 = \frac{3(N+1)}{10} \text{th item's value.}$$

$$= \frac{3(19+1)}{10} = \frac{60}{10} = 6 \text{ th items value.}$$

Value of 6th item $= 17$

$D_7 = \frac{7(N+1)}{10}$ th item's value.

$= \frac{7(19+1)}{10}$ th item's value $= \frac{140}{10}$ th items value.

$= 14$ th item's value which is 35.

$P_{35} = \frac{35(N+1)}{100}$ th item's value

$= \frac{35(19+1)}{100} = \frac{700}{100} = 7 \Rightarrow 18.$

$P_{80} = \frac{80(N+1)}{100}$ th item's value

$= \frac{80\,(19+1)}{100} = \frac{1600}{100} = 16 \Rightarrow 37.$

DISCRETE SERIES

Illustration No.2

From the following series calculate Q_1, Q_3, D_6, P_{47}:

Variables:	5	10	15	20	25	30	35	40
Frequency:	16	18	22	21	24	14	11	9

Solution

X	F	cf
5	16	16
10	18	34
15	22	56
20	21	77
25	24	101
30	14	115
35	11	126
40	9	135

$Q_1 = \frac{N+1}{4}$ th item's value

$= \frac{135+1}{4} = \frac{136}{4} = 34 \Rightarrow 10.$

$Q_3 = \frac{3(N+1)}{4}$ th item's value.

$$=\frac{3(135+1)}{4}=\frac{3\text{ x }136}{4}=102\ \Rightarrow 30$$

$$D_6=\frac{6(N+1)}{10}\text{ th item's value}$$

$$=\frac{6\ (135+1)}{10}=\frac{6\text{x}136}{10}=\frac{816}{10}=81.6\Rightarrow 25$$

$$P_{47}=\frac{47(N+1)}{100}\text{th item's value}$$

$$=\frac{47(135+1)}{100}=\frac{47\text{x}136}{100}$$

$$=\frac{6392}{100}=63.92\Rightarrow 20.$$

Class Interval Series

Illustration No.3

Calculate median, quartiles one and three, 3rd decile and 8th decile and 53rd percentile.

C.I.	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	16	14	23	17	7	3

Solution

C.I.	f	cf
0-10	16	16
10-20	14	30
20-30	23	53
30-40	17	70
40-50	7	77
50-60	3	80

$$Q_1=\frac{N}{4}\text{ th item's value}$$

$$=\frac{80}{4}\text{th item's value.}$$

$$=20\text{ th item's value which lies in (C.I. 10-20)}$$

By Applying Formula

$$Q_1=l_1+\frac{\frac{N}{4}-C}{f}\times i\text{, we get}$$

$$= 10 + \frac{20-16}{14} \times 10$$

$$= 10 + \frac{4 \times 10}{14}$$

$$= 10 + \frac{40}{14}$$

$$= 10 + 2.86$$

$$= 12.86 \text{ Ans.}$$

Q_2 (median) $= \frac{2N}{4}$th item's value

$= \frac{160}{4}$ th item's value

= 40 th item's value which lies in (C.I. 20-30)

By Applying Formula

$$Q_2 = l_1 + \frac{\frac{2N}{4} - C}{f} \times i \text{, we get}$$

$$= 20 + \frac{40-30}{23} \times 10$$

$$= 20 + \frac{10 \times 10}{23}$$

$$= 20 + \frac{100}{23}$$

$$= 20 + 4.35$$

$$= 24.35 \text{ Ans.}$$

$Q_3 = \frac{3N}{4}$ th item's value

$= \frac{240}{4}$ th item's value

= 60 th item's value which lies in (C.I. 30-40)

By Applying Formula

$$Q_3 = l_1 + \frac{\frac{3N}{4} - C}{f} \times i \text{, we get}$$

$$Q_3 = 30 + \frac{60-53}{17} \times 10$$

$$= 30 + \frac{7\times10}{17} = 30 + \frac{70}{17} = 30 + 4.12$$

$$= 34.12 \text{ Ans.}$$

Decile

D_3 lies in C.I.(10-20)
$$= l_1 + \frac{\frac{3\times N}{10} - C}{f} \times i$$

$$= 10 + \frac{\frac{3\times80}{10} - 16}{14} \times 10$$

$$= 10 + \frac{24-16}{14} \times 10$$

$$= 10 + \frac{80}{14}$$

$$= 10 + 5.71 = 15.71$$

$$= 15.71 \text{ Ans.}$$

D_8 lies in C.I. (30-40)
$$= l_1 + \frac{\frac{8\times N}{10} - C}{f} \times i$$

$$= 30 + \frac{\frac{8\times80}{10} - 53}{17} \times 10$$

$$= 30 + \frac{64-53}{17} \times 10$$

$$= 30 + \frac{110}{17}$$

$$= 30 + 6.47$$

$$= 36.47 \text{ Ans.}$$

Percentiles

P_{53} lies in C.I. (20-30)
$$= l_1 + \frac{\frac{53\ N}{100} - C}{f} \times i$$

$$= 20 + \frac{\frac{53x80}{100} - 30}{23} x10$$

$$= 20 + \frac{\frac{4240}{100} - 30}{23} x\ 10$$

$$= 20 + \frac{\frac{424-300}{10}}{23} x\ 10$$

$$= 20 + \frac{124}{23}$$

$$= 20 + 5.4$$

$$= 25.4 \text{ Ans.}$$

Illustration No.4

Calculate median, Q_1, Q_3, from the following data:

Marks: (C.I.)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students: (frequency)	5	8	11	15	7	3	1

Solution:

C.I.	f	cf	Marks Less than	No. of students
0-10	5	5	10	5
10-20	8	13	20	13
20-30	11	24	30	24
30-40	15	39	40	39
40-50	7	46	50	46
50-60	3	49	60	49
60-70	1	50	70	50

Verification by applying formulae

$\text{Med} = \frac{N}{2} = \frac{50}{2} = 25\text{th}$ item which lies in C.I. 30-40.

Applying formula

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 30 + \frac{\frac{50}{2} - 24}{11} \times 10$$

$$= 30 + \frac{25\text{-}24}{11} \times 10$$

$$= 30 + \frac{10}{11} = 30 + .91$$

$$= 30.91$$

$Q_1 = \frac{N}{4}$ th item's value

$= \frac{50}{4}$ th item's value

= 12.5 th items value which lies in C.I. (10-20)

Applying formula

$$= l_1 + \frac{\frac{N}{4} - C}{f} \times i$$

$$= 10 + \frac{\frac{50}{4} - 5}{8} \times 10$$

$$= 10 + \frac{12.5 - 5}{8} \times 10 \qquad = 10 + \frac{7.5}{8} \times 10$$

$$= 10 + 9.375$$

= 19.375 Ans.

$Q_3 = \frac{3 \times N}{4}$ th item's value

$= \frac{3 \times 50}{4}$ th items's value

= 37.5 th items's value which lies in C.I. 30-40

Applying formula

$$= l_1 + \frac{\frac{3N}{4} - C}{f} \times i$$

$$= 30 + \frac{\frac{3 \times 50}{4} - 24}{15} \times 10$$

$$= 30 + \frac{13.5}{15} \times 10$$

$$= 30 + 9$$

$$= 39$$ Ans.

Illustration No. 5

Calculate the Quartile 1 and 3 from the following data

Marks(below)	10	20	30	40	50	60	70	80
No. of students	15	35	60	84	96	127	198	250

Solution

First change the cumulative series into exclusive class interval series

C.I.	f	cf
0-10	15	15
10-20	20	35
20-30	25	60
30-40	24	84
40-50	12	96
50-60	31	127
60-70	71	198
70-80	52	250

$Q_1 \quad = \frac{N}{4}$ th item's value

$= \frac{250}{4}$ th item's value

$= 62.5$ th item's value which lies in (C.I. 30 - 40)

Applying formula

$$Q_1 = l_1 + \frac{\frac{N}{4} - C}{f} \times i$$

$$= 30 + \frac{\frac{250}{4} - 60}{24} \times 10$$

$$= 30 + \frac{62.5-60}{24} \times 10$$

$$= 30 + \frac{2.5}{24} \times 10$$

$$= 30 + 1.04$$

$$= 31.04 \text{ Ans.}$$

$$Q_3 = \frac{3N}{4} \text{ th item's value}$$

$$= \frac{3 \times 250}{4} \text{th items value}$$

$$= \frac{750}{4} \text{ th item's value}$$

$$= 187.5\text{th} \quad \text{item's value which lies in (C.I. 60-70)}$$

Applying formula

$$Q_3 = l_1 + \frac{\frac{3N}{4} - C}{f} \times i$$

$$= 60 + \frac{\frac{250 \times 3}{4} - 127}{71} \times 10$$

$$= 60 + \frac{187.5 - 127}{71} \times 10$$

$$= 60 + \frac{60.5}{71} \times 10$$

$$= 60 + 8.52$$

$$= 68.52 \text{ Ans.}$$

QUESTIONS

1. Find out the median and third quartile from the following data:

Height in Inches	58	59	60	61	62	63	64	65	66
No. of persons:	2	3	6	15	10	5	4	3	1

(Hint: Calculate cf and use N + 1/ 2 for M and 3x (N + 1)/ 4 for Q3)

2. Calculate the median Q3, D7, and P90 using data given below:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students:	3	10	17	7	6	4	2	1

3. Calculate the median and quartiles for the following:

Marks(below):	10	20	30	40	50	60	70	80
No. of students:	15	35	60	84	96	127	198	250

4. Calculate the median, first quartile and third decile from the following data:

Weekly wages(Rs)	58,	59,	60,	61,	62,	63,	64,	65,	66
No. of workers	2	3	6	15	10	5	4	3	1

5. The following are the marks of 50 students in statistics. Find Median Q_1 and Q3

Marks:	5-15	15-25	25-35	35-45	45-55
No.. of students:	8	12	15	9	6

6. Calculate the value of median, 1st Quartile, and 3rd Quartile from the following data·

Age:	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons:	50	70	100	180	150	120	70	60

7. The percentage marks obtained by 100 students in Economics in an examination are given below. Compute the median, 1st and 3rd quartiles:

Marks:	30-35	35-40	40-45	45-50	50-55	55-60	60-65
No. of students:	14	16	18	23	18	8	3

8. Calculate 6th decile, 65th percentile and 78th percentile from the following data:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	: 5	7	8	12	28	22	10	8

9. Find the values of quartiles, 7th decile and 80th percentile from the following data:

C.I.	0-10	10-20	20-30	30-40	40-50	50-60
	20	38	46	35	20	10

10. Find out the value of quartiles from the following:

Marks below:	10	10-20	20-30	30-40	40-50	50 and above
	8	10	22	25	10	5

11. Ascertain the values of quartiles, 6th decile and 35th percentile from the following data:

Marks less than:	80	70	60	50	40	30	20	10
No. of students:	100	90	80	60	32	20	13	5

5

Measures of Variability

We have seen that the statistical series differ from one another in several ways. There may be series having identical average but even then they may not be similar in many other aspects. No doubt, averages simplify the data to a large extent but they do not give the whole picture of the data.To know the difference in various series further analysis is necessary.

Let us take the example of per capita income of a country. It tells us only the level of economic development in a country, but does not tell us as to how the income is distributed among the people of the country. Similarly, by the average number of goals by two teams in league matches, we cannot know the consistent performance of two teams. We must find out the variations of goals by each team, in the pool or league matches.

Let us take another case of three series where the mean in all the cases is 10:

1 Series:	7	8	9	10	11	12	13
2 Series:	10	10	10	10	10	10	10
3 Series:	10	1	7	13	18	15	6

In the above three series, mean is 10 but all of them differ in respect of variability. In one case all the values of the series coincide with the mean. In the first series, all the values are nearing mean but in the third series, almost all the values deviate from the mean. Thus, average fail to give us correct picture.It is for this reason that we should study further by examing the variations in the series.

Definition of Variability:

According to Prof. Spiegel "*The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data*".

Requisites for a measure of Dispersion

1. Should be rigidly defined.
2. Be easy to calculate and understand.

3. Based on all the observations.

4. Can be mathematically calculated.

Various Measures of dispersion

Following are some of the commonly used methods, for measuring variations or dispersion:

a) Range
b) Quartile Deviation
c) Mean Deviation
d) Standard Deviation
e) Co-efficient of Variation
f) Lorenz Curve.

I. RANGE

It is the simplest measure of dispersion. It is defined as the difference between the maximum value and the minimum value of the items in a series. *The difference between the greatest and the least values of the variable in a distribution is called range.*

It is very easy to locate range. By looking at the two extremes of the series, we can calculate or locate the range. For lexample, weekly pocket expenditure of 8 students is rupees 12, 14, 20, 35, 40, 37, 50. The smallest pocket expenditure is Rs.12 and the highest is Rs. 50. Thus the range is 50-12 = Rs. 38.

In the calculation of range the following formula is used:

$$\text{Range} = L - S$$

Here "L" stands for the largest value of the class and "S" stands for the smallest value of the class.

Range takes into consideration only the two extreme values of the distribution viz. the highest and the lowest. It ignores other values of the variable. The relative measure of range is defined as under:-

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$

$$= \frac{\text{Highest item-smallest item}}{\text{Highest item + smallest item}}$$

in the above example,

$$\text{Co-efficient of Range} = \frac{50-12}{50+12} = \frac{38}{62}$$

$$= 0.61$$

Calculation of Co-efficient of Range

Individual Series

Illustration No.1

Find range and co-efficient of range from the following data:

25, 27, 29, 30, 31, 38, 48, 35, 50

Solution

$R = L - S = 50 - 25 = 25$

Co-efficient of Range: $\dfrac{L-S}{L+S} = \dfrac{50-25}{50+25} = \dfrac{25}{75} = 0.33$

Discrete Series

Illustration No.2

Find out the range and co-efficient of range of the following distribution:

x	6	7	8	9	10	11	12	15
f	10	15	17	80	25	20	10	4

Solution

Here the maximun value of x is 15 and the smallest is 6, thus,

Range $= L - S$

$= 15 - 6 = 9$

Co-efficient of Range $= \dfrac{L-S}{L+S} = \dfrac{15-6}{15+6} = \dfrac{9}{21} = 0.42$

Illustration No.3

Location of Range in the class-interval series:

Find out the range and its co-efficient from the following series:

C.I.	5-10	10-15	15-20	20-25	25-30	30-35	35-40
	3	8	7	12	9	8	6

Solution

Here $L = 40$ and $S = 5$

$R = L - S = 40 - 5 = 35$

Co-efficient of Range $= \dfrac{L-S}{L+S} = \dfrac{40-5}{40+5} = \dfrac{35}{45} = 0.78$

Merits and Demerits of Range

Merits:

1. It is the simplest measure of dispersion.

2. It is easily computable.

3. It is helpful in further mathematical calculations and helpful in quality control and weather forecast reporting.

4. We use it in our daily life.

Demerits:

1. It is not based on all items of the distribution.
2. It is considered the poorest method of measuring variations.
3. It is not affected by the middle value of the variables.
4. Its calculation is difficult in open-end series.
5. It is highly affected by the extreme values.

2. Quartile Deviation

Quartile deviation is an absolute measure of dispersion. The effect of extreme values on range can be avoided by adopting the measure of *quartile range. It is defined as half of the difference betweenthe third and first quartile, whereas the difference between the third and first quartile is called interquartile range. Quartile deviation is also called semi-interquartile range*. Hence

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

Where Q_1 is the lower quartile (first quartile) and Q_3 is the upper quartile (third quartile).

Inter-quartile range $= Q_3 - Q_1$

Co-efficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

Calculation of Inter-quartile Range and Quartile Deviation In Individual Series

Illustration No.4

Find out the inter-quartile range and its co-efficient from the following data:

Monthly pocket money of students are:

18 19 21 22 23 27 30 31 37 45 52

$$Q_1 = \frac{N+1}{4} \text{ th item}$$

X
18
19
21
22
23
27
30
31
37
45
52

$$= \frac{11+1}{4} = 3 \text{ rd item} = 21$$

$$Q = \frac{3(N+1)}{4} \text{th item}$$

$$= \frac{3\,(11+1)}{4} = \frac{36}{4} = 9\text{th item} = 37$$

$$\text{Quartile Range} = Q_3 - Q_1 = 37 - 21 = 16$$

Semi-inter Quartile Range

$$= \frac{Q_3 - Q_1}{2} = \frac{37-21}{2} = \frac{16}{2} = 8$$

$$\text{Co-efficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{37-21}{37+21} = \frac{16}{58} = 0.28 \text{ approx.}$$

Discrete Series

Illustration No.5

Find out the Quartile Deviation, inter quartile range and co-efficient of quartile deviation for the following distribution:

x	6	7	8	9	10	11	12	13
f	10	15	17	30	25	20	10	4

Solution

x	f	cf
6	10	10
7	15	25
8	17	42
9	30	72
10	25	97
11	20	117
12	10	127
13	4	131

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{th item} = \frac{131+1}{4} = \frac{132}{4} = 33\text{rd item}$$

33rd item = 8

$$Q_3 = \text{size } \frac{3(N+1)}{4} \text{th item.} = \frac{3\text{x}132}{4} = 99\text{th item}$$

99th item = 11

Therefore, quartile deviation $= \frac{Q_3-Q_1}{2} = \frac{11-8}{2} = \frac{3}{2} = 1.5$

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{11-8}{11+8} = \frac{3}{19} = 0.16$$

Class Interval Series

Illustration No.6

Find out the quartile deviation and co-efficient of quartile deviation for the following distribution:

C.I.	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
f	6	6	14	10	3	1	2	2

Solution

C.I.	f	cf
5-10	6	6
10-15	6	12
15-20	14	26
20-25	10	36
25-30	3	39
30-35	1	40
35-40	2	42
40-45	2	44

$$Q_1 = \frac{N}{4} = \frac{44}{4} = 11\text{th item}$$

The 11th item lies in C.I. 10-15 so the lower quartile class is 10-15

Apply formula

$$Q_1 = l_1 + \frac{\frac{N}{4} - C}{f} \times i$$

$$= 10 + \frac{\frac{44}{4} - 6}{6} \times 5$$

$$= 10 + \frac{11-6}{6} \times 5$$

$$= 10 + \frac{5}{6} \times 5$$

$$= 10 + \frac{25}{6}$$

$$= 10 + 4.167$$

$Q_1 = 14.167$

Now $Q_3 = \frac{3 \times N}{4} = \frac{3 \times 44}{4}$ th item

= size of 33 rd item

The 33rd item lies in the C.I. 20-25, the upper quartile class is 20-25

Applying formula for Q_3

$$= 20 + \frac{3 \times 44}{4} - \frac{26}{10} \times 5$$

$$= 20 + \frac{33-26}{10} \times 5$$

$$= 20 + 3.5 = 23.5$$

Therefore $Q_3 = 23.5$

Hence quartile deviation $= \frac{Q_3 - Q_1}{2}$

$$= \frac{23.5 - 14.16}{2}$$

$$= \frac{9.34}{2} = 4.67$$

Co-efficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{23.5 - 14.16}{23.5 + 14.16} = \frac{9.34}{37.66} = 0.248$$

Illustration No.7

Find out the quartile deviation and co-efficient of quartile deviation from the following data:

Less than 10	20	30	40	50	60	70
f 5	8	15	20	30	33	35

Solution

First of all the cumulative frequency distribution is to be changed into ordinary frequency distribution to know the clas interval.

CI	f	cf
0-10	5	5
10-20	3	8
20-30	7	15
30-40	5	20
40-50	10	30
50-60	3	33
60-70	2	35

$$Q_1 = \text{size} \frac{N}{4} = \frac{cf}{4} \text{ item} = \frac{35}{4} = 8.75 \text{ th item}$$

8.75th item lies in the C.I. of 20-30

By applying formula of Q_1

$$= 20 + \frac{8.75 - 8}{7} \times 10$$

$$= 20 + \frac{7.5}{7} = 21.07$$

$Q_1 = 21.07$

$$\text{Now } Q_3 = \text{size } \frac{3 \times N}{4} = \frac{3 \times cf}{4} = \frac{3 \times 35}{4} = 26.25 \text{ th item}$$

= 26.25 th item lies in C.I. of 40-50

By applying formula of Q_3

$$= 40 + \frac{26.25 - 20}{10} \quad \times 10$$

$$= 40 + \frac{6.25 \times 10}{10}$$

$= 46.25$

$Q_3 = 46.25$

Quartile Deviation $= Q_3 - Q_1 = 46.25 - 21.07 = 25.18$

$$\text{Co-efficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{46.25 - 21.07}{46.25 + 21.07}$$

$$= \frac{25.18}{67.32}$$

$$= 0.374$$

Merits and Demerits of Quartile Deviation

Merits:

1. It can be easily calculated.
2. It is easy to understand.
3. It is rigidly defined.
4. It is not affected by the extreme valued of items.

Demerits

1. It ignores many items.
2. It is affected by fluctuations in the samples.
3. Further algebraic treatment is not possible.
4. It is not a fine measure of dispersion.

MEAN DEVIATION

Mean deviation of a series *is the arithmetic average of the deviations (in absolute terms i.e. signs of deviations are ignored) of various items from a measure of central tendency*. Generallly, mean deviation is calculated from mean or median. It shows the scatter or spread of variable around the average.

Since, we have to measure the scatter, we do not bother if the observed values are more or less than the central tendency. In other words, we ignore the plus and minus signs of deviation, sum them up and find the arithmetic mean of deviation by dividing the sum by the number of items.

Calculation of Mean Deviation

Individual Series

While calculating mean deviation, we proceed in the following manner:

a) First of all calculate mean or median from the series.

b) Take deviations of all given values of the variable from the mean or median. If we take deviations from mean we represent them as (dx). If we take it from median, we call it (dm).

c) Then add all the deviations ignoring their plus and minus signs and then divide this by the number of the values of the variable.

d) The formulae are: When deviations are taken from mean

i) Mean Deviation (M.D.) $= \frac{\Sigma\,|dx|}{N}$

ii) Co-efficient of mean deviation $= \frac{M.D.}{Mean}$

When deviations are taken from median

iii) Mean Deviation $= \frac{\Sigma\,|dm|}{N}$

iv) Co-efficient of Mean Deviation $= \frac{M.D.}{Mean}$

Illustration No.8

Calculate mean deviation form the following. Also calculate the co-efficients of mean deviation.

Take deviations from mean and median.

15 17 19 25 30 35 48

Solution :when deviations are taken from mean.

x	dx
15	12
17	10
19	8
25	2
30	3
35	8
48	21
$\Sigma x = 189$	$\Sigma\|dx\| = 64$

Mean Deviation

i) Mean $(\overline{X}) = \frac{\Sigma x}{N} = \frac{189}{7} = 27$

ii) Mean Deviation $= \frac{\Sigma \mid dx \mid}{N}$

$$= \frac{64}{7} = 9.14$$

iii) Co-efficient of Mean Deviation

$$= \frac{M.D.}{Mean} = \frac{9.14}{27} = 0.33$$

When deviations are taken from Median

x	dm
15	10
17	8
19	6
25	0
30	5
35	10
48	23
$\Sigma x = 189$	$\Sigma \mid dm \mid = 62$

i) Median $= \frac{N+1}{2}$th item

$$= \frac{7+1}{2} = \text{4th item} = 25$$

ii) Mean Deviation $\frac{\Sigma . \mid dm \mid}{N}$

$$= \frac{62}{7} = 8.85$$

iii) Co-efficient of Mean Deviation =

$$\frac{M.D.}{Median} = \frac{8.85}{25} = 0.36$$

Discrete Series

Steps in Solution

1. Find out Mean or Median from the series.

2. Take deviations from the Mean or Median. Ignore plus and minus signs and denote it as $|dx|$ in case of mean and $|dm|$ in case of median.

3. Multiply $|dx|$ or $|dm|$ by respective frequency and denote it as $f|dx|$ or $f|dm|$.

4. Add up $f|dx|$ or $f|dm|$.

5. Apply formula: Mean Deviation $= \dfrac{\Sigma f|dx|}{N}$

or Mean Deviation $= \dfrac{\Sigma f|dm|}{N}$

6. Co-efficient formulae as in illustration no.8

Illustration No.9

Calculate Mean Deviation, Median Deviation and Coeffcient of Mean Deviations :

Size	10	11	12	13	14
F:	3	12	18	12	3

Solution

Mean Deviation

x	f	fx	|dx|	f|dx|
10	3	30	2	6
11	12	132	1	12
12	18	216	0	0
13	12	156	1	12
14	3	42	2	6
	$\Sigma f = 48$	$\Sigma fx = 576$		$\Sigma f|dx| = 36$

$$\text{Mean } \overline{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{576}{48} = 12$$

$$\text{Mean Deviation} = \frac{\Sigma f|dx|}{\Sigma f} = \frac{36}{48} = 0.75$$

$$\text{Co-efficient of Mean Deviation} = \frac{\text{M.D.}}{\text{Mean}} = \frac{0.75}{12}$$

$$= 0.0625$$

MEDIAN DEVIATION

x	f	cf	|dm|	f|dm|
10	3	3	2	6
11	12	15	1	12
12	18	33	0	0
13	12	45	1	12
14	3	48	2	6
	$\Sigma f = 48$			$\Sigma f \mid dm \mid = 36$

$$\text{Median} = \frac{N+1}{2} = \frac{48+1}{2} = 24.5\text{th item}$$

$$= 12$$

$$\text{Median Deviation} = \frac{\Sigma f \mid dm \mid}{\Sigma f} = \frac{36}{48} = 0.75$$

Co-efficient of Median Deviation =

$$\frac{\text{M.D.}}{\text{Median}} = \frac{0.75}{12} = 0.0625 \text{ Ans.}$$

Class-Interval Series

Mean deviation and Co-efficient of Mean Deviation

Illustration No.10

Calculate mean deviation and co-efficient of mean deviation from the following:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
frequency	18	16	15	12	10	5	2	2

Solution

Take mid-values of class intervals and compute mean with them. From mean, take deviation, multiply the deviation taken from mean by frequency. add up f |dx | column and denote it as Σf |dx | and also add up frequency column i.e. obtain Σf. Mean deviation from mean can be obtained by the following formula-

$$\text{Mean deviation from Mean} = \frac{\Sigma f \mid dx \mid}{\Sigma f}$$

$$\text{Co-efficient of mean deviation} = \frac{\text{Mean deviation}}{\overline{X}} = \frac{\text{M.D.}}{\text{Mean}}$$

Table No.1

A = 35

C.I	Mid-value	f	Step deviation from A = 35	Product fdx
0-10	5	18	-30	-240
10-20	15	16	-20	-320
20-30	25	15	-10	-150
30-40	35	12	0	0
40-50	45	10	10	100
50-60	55	5	20	100
60-70	65	2	30	60
70-80	75	2	40	80
		$\Sigma f = 80$		$\Sigma fdx = -370$

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f}$$

$$= 35 - \frac{370}{80} = 35 - 4.62$$

$$\overline{X} = 30.38$$

Table No.2

f	Mid value	Deviation from Mean $\|dx\|$ 30.38	Product f $\|dx\|$
18	5	25.38	456.84
16	15	15.38	246.08
15	25	5.38	80.70
12	35	4.62	55.44
10	45	14.62	146.20
5	55	24.62	123.10
2	65	34.62	69.24
2	75	44.62	89.24
$\Sigma f = 80$			$\Sigma f\|dx\| = 1266.84$

$$\text{Mean deviation} = \frac{\Sigma f\,|dx|}{\Sigma f}$$

$$= \frac{1266.84}{80} = 15.83$$

$$\text{Co-efficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}}$$

$$= \frac{15.83}{30.38}$$

$$= 0.521$$

Illustration No.11

Calculate median deviation and co-efficient of median deviation.

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
frequency	18	16	15	12	10	5	2	2

Solution

C.I.	M.V	f	cf	$\lvert dm \rvert$ from 24	$f\lvert dm \rvert$
0-10	5	18	18	19	342
10-20	15	16	34	9	144
20-30	25	15	49	1	15
30-40	35	12	61	11	132
40-50	45	10	71	21	210
50-60	55	5	76	31	155
60-70	65	2	78	41	82
70-80	75	2	80	51	102
		$\Sigma f = 80$			$\Sigma f\lvert dm \rvert = 1182$

$$\text{Median} \frac{N}{2} \text{ item} = \frac{80}{2} = 40\text{th item}$$

40th item is in C.I. = 20-30

Applying formula-

$$\text{Median} = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$20 + \frac{\frac{80}{2} - 34}{15} \times 10$$

$$= 20 + \frac{6}{15} \times 10 = 24$$

$$\text{Median deviation} = \frac{\Sigma f\,|dm|}{\Sigma f}$$

$$= \frac{1182}{80} = 14.77$$

$$\text{Co-efficient of median deviation} = \frac{\text{M.D}}{\text{Median}}$$

$$= \frac{14.77}{12} = 0.615$$

Illustration No.12

Compute the mean and median deviations and their co-efficients

x	140-150	150-160	160-170	170-180	180-190	190-200
f	4	6	10	18	9	3

Solution

A = 165 i = 10 | dx | from 171.2 | dm| from 172.8

C.I.	M.V.	f	dx	fdx	|dx|	f|dx|	|dm|	f|dm|
140-150	145	4	-2	-8	26.2	104.8	27.8	111.2
150-160	155	6	-1	-6	16.2	97.2	17.8	106.8
160-170	165	10	0	0	6.2	62.0	7.8	78.0
170-180	175	18	+1	+18	3.8	68.4	2.2	39.6
180-190	185	9	+2	+18	13.8	124.2	12.2	109.8
190-200	195	3	+3	+9	23.8	71.4	22.2	66.6
		Σf = 50		Σfdx =+31		f |dx| = 528.0		Σf|dm =512.0

$$\overline{X} = A + \frac{\Sigma fdx}{\Sigma f} \times i$$

$$= 165 + \frac{31}{50} \times 10$$

$$= 165 + \frac{31}{5}$$

$$= 165 + 6.2$$

$$= 171.2$$

$$\text{Median} = \frac{N}{2} \text{ size of item}$$

$$= \frac{50}{2} = 25 \text{ th item}$$

which lies in Class interval 170– 180

$$\text{Median} = l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

Mean Deviation

$$= \frac{\Sigma|dx|}{\Sigma f}$$

$$= \frac{528}{50}$$

$$= 10.56$$

$$= 170 + 25\text{–}20/18\text{x}10$$

$$= 170 + \frac{50}{18} \quad = 172.8$$

Median deviation

$$= \frac{\Sigma|dm|}{\Sigma f} = \frac{512}{50} = 10.24$$

Co-efficient of Mean deviation =

$$\frac{\text{Mean deviation}}{\text{Mean}}$$

$$= \frac{10.56}{171.2}$$

$$= 0.061$$

Co-cfficient of Median deviation =

$$\frac{\text{Median deviation}}{\text{Median}}$$

$$= \frac{10.24}{172.80}$$

$$= 0.059$$

Illustration No.13

Compute the mean deviation and the coefficient of mean deviation from the following data:

C.I.	0-20	20-40	40-60	60-80	80-100	100-120
frequency	5	50	84	32	10	6

Solution

A = 50 i = 20 |dx| from 51.06

x	MV	f	dx	dx′	fdx′	|dx|	f|dx|
0-20	10	5	-40	-2	-10	41.06	205.30
20-40	30	50	-20	-1	-50	21.06	1053.00
40-60	50	84	0	0	0	1.06	89.04
60-80	70	32	+20	+1	+32	18.94	606.08
80-100	90	10	+40	+2	+20	38.94	389.40
100-120	110	6	+60	+3	+18	58.99	353.64
		Σf = 187			Σfdx = 10		Σf |dx| = 2696.46

$$\overline{X} = A \pm \frac{\Sigma fdx}{\Sigma f} \quad x\ i$$

$$= 50 + \frac{10}{187} \times 20$$

$$= 50 + \frac{200}{187} \quad = 51.06$$

Mean deviation

$$= \frac{\Sigma f\,|dx|}{\Sigma f}$$

$$= \frac{2696.46}{187}$$

$$= 14.41$$

Co efficient of Mean deviation

$$= \frac{\text{Mean deviation}}{\text{Mean}}$$

$$= \frac{14.41}{51.06}$$

$$= 0.282$$

Illustration No.14

Calculate the median deviation and coefficient of median deviation from the following data:

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	4	6	8	10	7	6	5	3	1

Solution

x	MV	f	cf	|dm| from 37	f|dm|
0-10	5	4	4	32	128
10-20	15	6	10	22	132
20-30	25	8	18	12	96
30-40	35	10	28	2	20
40-50	45	7	35	8	56
50-60	55	6	41	18	108
60-70	65	5	46	28	140
70-80	75	3	49	38	114
80-90	85	1	50	48	48
		Σf = 50			Σf|dm| = 734

Median = N/2th size of items.

= 50/2 = 25th item. It lies in C.I. 30-40

$$= l_1 + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 30 + \frac{25 - 18}{10} \times 10 = 30 + 7 = 37$$

$$\text{Median deviation} = \frac{\Sigma f \mid dm \mid}{\Sigma f} = \frac{734}{50} = 14.6$$

Coefficient of Median deviation =

$$\frac{\text{M.D.}}{\text{Median}} = \frac{14.6}{37} = 0.39$$

Merits & Demerits of Mean Deviation

Merits:

1. Defined rigidly.
2. Easily understandable.
3. Easy to calculate.
4. Calculations based on all items.
5. Less affected by the extreme values.
6. Better measure of dispersion.

Demerits

1. Not helpful in further analysis because further mathematical treatment is limited.
2. Ignores plus and minus signs.
3. Not so popular as standard deviation.

STANDARD DEVIATION

Standard deviation is the square root of the arithmetic mean of the squares of deviations of the items from their arithmetic mean.

Calculation of the Standard Deviation

Individual Series

Steps in calculation of the S.D. from individual series are:-

i) First, find out the $\overline{X}$ (Mean)

ii) Secondly, take deviations from the mean and find out the squares of these deviations and denote them as dx^2.

iii) Thirdly, total up the squares of these deviations and denote them as Σdx^2

iv) Fourthly, divide the sum as taken above by number of items.

v) Find the square root of the result obtained in step (iv).

vi) The formula of standard deviation is: standard deviation or

$$S.D. = \sigma\sqrt{\frac{dx^2}{N}}$$

Illustration No.15

Find standard deviation from the following:-

3, 4, 6, 7, 10

Solution

x	dx from 6	dx^2
3	-3	9
4	-2	4
6	0	0
7	1	1
10	4	16
$\Sigma x = 30$		$\Sigma dx^2 = 30$

$$\overline{X} = \frac{\Sigma x}{N}$$

$$= \frac{30}{5} = 6$$

$$\sigma = \sqrt{\frac{\Sigma dx^2}{N}}$$

$$= \sqrt{\frac{30}{5}} = \sqrt{6}$$

$$= 2.44$$

Discrete and Continuous series

For computing the standard deviation of the discrete series, we can use either of three methods viz. (i) direct method (ii) short-cut method and (iii) step deviation method.

Direct Method

In this method, following procedure is adopted for calculating standard deviation:-

i) First find out the arithmetic mean.

ii) Take deviations from the mean and square these deviations.

iii) Multiply the squares of deviations by their respective frequencies.

iv) Then total up fdx^2 and divide it by the total of frequencies.

v) Compute the square root of the result obtained in step no.(iv) above.

vi) The formula thus becomes

$$S.D. = \sqrt{\frac{\Sigma f\,(dx)^2}{\Sigma f}}$$

Illustration No.16

Calculate the standard deviation of the following data:-

x	6	7	8	9	10	11	12
f	7	6	9	13	8	5	2

Solution

x	f	fx	Deviation from Mean i.e.8.64	Square of deviations	Product $f(dx)^2$
6	7	42	-2.64	6.97	48.79
7	6	42	-1.64	2.69	16.14
8	9	72	-0.64	0.41	3.69
9	13	117	0.36	0.13	1.69
10	8	80	1.36	1.85	14.80
11	5	55	2.36	5.57	27.85
12	2	24	3.36	11.29	22.58
	$\Sigma f = 50$	$\Sigma fx = 432$			$\Sigma f(dx)^2 =$ 135.54

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{432}{50} = 8.64$$

$$S.D. = \sqrt{\frac{\Sigma f(dx)^2}{\Sigma f}}$$

$$= \sqrt{\frac{135.54}{50}}$$

$$= \sqrt{2.7108}$$

S.D. = 1.65

Illustration No.17

Short-cut Method

Find standard deviation from the following data:

x:	10	12	14	16	18	20	22
f:	3	5	9	16	8	7	2

In the short-cut method, we take deviation from assumed mean and denote them by dx. This dx are multiplied by thier respective frequencies. There after, fdx is multiplied by dx to obtain fdx^2. After applying the following formula, we can find out the standard deviation:-

$$\text{Formula for S.D} = \sqrt{\frac{\Sigma fdx^2}{\Sigma f} - \left(\frac{\Sigma fdx}{\Sigma f}\right)^2}$$

Solution

x	f	dx	fdx	fdx^2
10	3	–4	–12	48
12	5	–2	–10	20
14	9	0	0	0
16	16	2	32	64
18	8	4	32	128
20	76	42	2	52
22	2	8	16	128
	$\Sigma f = 50$		$\Sigma fdx = 100$	$\Sigma fdx^2 = 640$

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx^2}{\Sigma f} - \left(\frac{\Sigma fdx}{\Sigma f}\right)^2}$$

$$\sigma = \sqrt{\frac{640}{50} - \left(\frac{100}{50}\right)^2}$$

$$= \sqrt{12.8 - 4.00} \quad = \sqrt{8.8} \quad = 2.96$$

Class Interval Series:

In such series, we first take mid-values of the class intervals. A mean is assumed and from that assumed mean, we take deviations. The deviations are multiplied by frequencies. The fdx is multiplied by dx. Then following formula is applied :

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx^2}{\Sigma f} - \left(\frac{\Sigma\, fdx}{\Sigma f}\right)^2}$$

Illustration No.18

Calculate the standard deviation from the following data:-

C.I.	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
frequency	7	12	15	18	20	16	8	4

Solution

AM = 55

C.I.	M.V.	f	dx	fdx	fdx^2
20-30	25	7	–30	–210	6300
30-40	35	12	–20	–240	4800
40-50	45	15	–10	–150	1500
50-60	55	18	0	0	0
60-70	65	20	10	200	2000
70-80	75	16	20	320	6400
80-90	85	8	30	240	7200
90-100	95	4	40	160	6400
		$\Sigma f = 100$		$\Sigma fdx = 320$	$\Sigma fdx^2 = 34600$

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx^2}{\Sigma f} - \left(\frac{\Sigma fdx}{\Sigma f.}\right)^2}$$

$$= \sqrt{\frac{34600}{100} - \left(\frac{320}{100}\right)^2}$$

$$= \sqrt{346 - 10.24}$$

$$= \sqrt{335.76}$$

$$\sigma = 18.32$$

STEP DEVIATION METHOD

In this method, mid-value is taken from Class Intervals. Mean is assumed. Deviations are taken from the mean. Common factor is taken out from dx and it it is termed as dx′. This dx′ is multiplied by the respective frequencies and is termed as fdx′. This fdx′ is multiplied by dx′ to make it fdx'^2. Now, the following formula is applied:-

$$\sigma = \text{S.D.} = \sqrt{\frac{\Sigma fdx'^2}{\Sigma f} - \left(\frac{\Sigma fdx'}{\Sigma f}\right)^2} \times i$$

Illustration No.19

Calculate the arithmetic mean and standard deviation from the following series:-

C.I.	5-15	15-25	25-35	35-45	45-55
Frequency	8	12	15	2	6

Solution

A = 30 i = 10

C.I.	M.V.	f	dx	dx'	fdx'	fdx'^2
5-15	10	8	−20	−2	−16	32
15-25	20	12	−10	−1	−12	12
25-35	30	15	0	0	0	0
35-45	40	9	10	1	9	9
45-55	50	6	20	2	12	24
		Σf = 50			Σfdx = −7	$\Sigma fdx'^2$ = 77

$$\text{S.D.} = \sqrt{\frac{\Sigma dx'^2}{\Sigma f} - \left(\frac{\Sigma fdx'}{\Sigma f}\right)^2} \times i$$

$$= \sqrt{\frac{77}{50} - \left(\frac{-7}{50}\right)^2} \times 10$$

$$= \sqrt{\frac{77}{50} - \frac{49}{2500}} \times 10$$

$$= \sqrt{\frac{3801}{2500}} \times 10 = \frac{61}{50} \times 10 = 12.2$$

Illustration No. 20

Calculate the mode and standard deviation from the following data:-

Wages, upto Rs	No. of workers
10	12
20	30
30	65
40	107
50	157
60	202
70	222
80	230

Solution

A = 45 i = 10

x	f	cf	M.V.	(x-45)dx	dx'	fdx'	fdx'^2
0-10	12	12	5	-40	-4	-48	192
10-20	18	30	15	-30	-3	-54	162
20-30	35	65	25	-20	-2	-70	140
30-40	42	107	35	-10	-1	-42	42
40-50	50	157	45	0	0	0	0
50-60	45	202	55	10	1	+45	45
60-70	20	22	65	20	2	+40	80
70-80	8	230	75	30	3	+24	72
	Σf = 230					Σfdx = -105	Σfdx^2 = 733

From observation we can say that mode lies in class interval 40-50.

$$\text{Mode} = l_1 + \frac{fm - c}{2fm - f_1 - f_2} \times i$$

$$= 40 + \frac{50-42}{2 \times 50 - 42 - 45} \times 10$$

$$= 40 + \frac{8}{13} \times 10$$

$$= 46.15$$

Standard Deviation

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx'^2}{\Sigma f} - \left(\frac{fdx'}{\Sigma f}\right)^2} \times i$$

$$= \sqrt{\frac{733}{230} - \left(\frac{105}{230}\right)^2} \times 10$$

$$= \sqrt{3.18 - \frac{11025}{52900}} \times 10$$

$$= \sqrt{3.18 - 0.20} \times 10$$

$$= \sqrt{2.98} \times 10$$

$=1.72 \times 10$

$=17.26$

Illustration No.21

Find the standard deviation of the height of 100 students:-

Height(in inches)	frequency
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Solution

A=67 i=3

C.I.	M.V.	dx	dx	f	dx'^2	fdx'	fdx'^2
60-62	61	-6	-2	5	4	-10	20
63-65	64	-3	-1	18	1	-18	18
66-68	67	0	0	42	0	0	0
69-71	70	+3	+1	27	1	27	27
72-74	73	+6	+2	8	4	16	32
				Σf =100		Σfdx =15	Σfdx^2 =97

$$S.D. = \sqrt{\frac{\Sigma fdx'^2}{\Sigma f} - \left(\frac{\Sigma fdx}{\Sigma f}\right)^2} \times i$$

$$= \sqrt{\frac{97}{100} - \left(\frac{15}{100}\right)^2} \times 3$$

$$= \sqrt{0.97 - 0.0225} \times 3$$

$$= 2.92$$

Merits and Demerits of Standard Deviation

Merits

1. Defined rigidly.
2. Further mathematical treatment possible.
3. Sampling fluctuations do not affect it.
4. By squaring deviations, we overcome the difficulty of mean deviation.

5. Based on the observations of all items.
6. Its value is very precise.

Demerits

1. It is difficult to calculate.
2. Extreme value unduly affect it by squaring deviations.
3. It is very difficult to understand.

Coefficient of Variation

The coefficient of variation is mostly used for the comparison of two or more than two distributions. The coefficient of standard deviation is defined as the quotient of the standard deviation and the mean. It is S.D. / Mean and the coefficient of variation is 100 times the coefficient of dispersion i.e. coefficient of variation (C.V).

$$\text{C.V.} = 100 \times \frac{\text{S.D.}}{\text{Mean}}$$

Coefficient of variation being a relative measure of dispersion, has certain advantages over standard deviation in certain situations. When we use this coefficient for comparison of two or more than two distributions, the coefficient of which is least, is said to be more consistent.

Illustration No.22

In a given distribution mean is 20, and S.D. = 16. Find the coefficient of variation

Solution

$$\text{C.V.} = \frac{\text{S.D.}}{\overline{X}} \times 100$$

$$= \frac{16}{20} \times 100 = 80\%$$

Illustration No.23

Calculate the variance of the number 2,7 and9.

$$\text{Mean} = 2 + 7 + 9 = \frac{18}{3} = 6$$

Solution

X	$x - \bar{x} = dx$	$(x-\bar{x})^2 = dx^2$
2	- 4	16
7	+1	1
9	+3	9
		26

$$\text{Variance} = \frac{\Sigma (x - x)^2}{N} = \frac{26}{3} = 8.76$$

Illustration No.24

The mean and standard deviations of the incomes of teachers of two schools A and B are given below:-

School	Mean Income (Rs)	Standard Deviation(Rs)
A	800	120
B	900	135

Compare the variability of incomes of teachers in the above two schools.

Solution

The variability of incomes of teachers in the A and B Schools can be compared by obtaining their co-efficient of variation.

$$\text{C.V. of school A} = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{120}{800} \times 100$$

$$= 15\%$$

$$\text{C.V. of school B} = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{135}{900} \times 100$$

$$= 15\ \%$$

In both the cases C.V. is equal.

Illustration No. 25

The co-efficients of variation of two series are 60% and 80%. The standard deviations in their cases are 20 and 16. Calculate arithmetic mean.

Solution

$$\text{In the first case C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$60 = \frac{20}{\overline{X}} \times 100$$

$$60\,\overline{X} = 2000$$

$$\text{or} \quad \overline{X} = \frac{2000}{60} = 33.3$$

In the second case:

$$80 = \frac{\sigma}{\overline{X}} \times 100 \qquad 80 = \frac{16}{\overline{X}} \times 100$$

$$\text{or } 80\,\overline{X} = 1600 \qquad \text{or } \overline{X} = 20$$

Thus arithmetic mean in the first case is 33.3 and in the second case it is 20.

Illustration No.26

Sarita and Mamta of Govt. Girls Senior Secondary School, Delhi obtained the following percentage of marks in 10 tests. The scholarship is to be awarded on the basis of consistency of performance of the candidate.

Find out who should get the scholarship:

Test	1	2	3	4	5	6	7	8	9	10
Sarita	44	80	71	48	52	72	68	56	60	64
Mamta	48	75	50	60	63	69	72	51	57	66

Solution

For finding out consistency we should know the C.V. in both the cases.

Computation of Coefficient of variation

Test	Sarita x	dx from 60	dx^2	Mamta Y	dy from 60	dy^2
1.	44	−16	256	48	−12	144
2.	80	20	400	75	15	225
3.	71	11	121	54	−6	36
4.	48	−12	144	60	0	0
5.	52	−8	64	63	3	9
6.	72	12	144	69	9	81
7.	68	8	64	72	12	144
8.	56	−4	16	51	−9	81
9.	60	0	0	57	−3	9
10.	64	4	16	66	6	36
	Σx $=615$	Σdx $=15$	Σdx^2 $=1225$	ΣY $=615$	Σdy $=15$	Σdy^2 $=765$

Sarita (X)

$$\overline{X} = \frac{\Sigma x}{\Sigma f} = \frac{615}{10} = 61.5$$

$$\sigma = \text{S.D.} = \sqrt{\frac{\Sigma dx^2}{\Sigma f} - \left(\frac{\Sigma dx}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{1225}{10} - \left(\frac{15}{10}\right)^2}$$

$$= \sqrt{122.5 - 2.25}$$

$$= \sqrt{120.25}$$

$$= 10.96$$

Mamta (Y)

$$\overline{Y} = \frac{615}{10} = 61.5$$

$$\text{S.D.} = \sqrt{\frac{\Sigma dy^2}{\Sigma f} - \left(\frac{\Sigma dy}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{765}{10} - \left(\frac{15}{10}\right)^2}$$

$$= \sqrt{76.5 - 2.25}$$

$$= \sqrt{74.2}$$

$$= 8.61$$

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{10.96}{61.5} \times 100$$

$$= 17.83$$

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{8.61}{61.5} \times 100$$

$$= 14.01$$

Since the C.V. of Mamta is low, she is entitled for the award.

Illustration No.27

In the C.B.S.E. examination in English, the mean marks of 150 students were 78 and the standard deviation was 80. In Economics, however,the mean marks of the group were 73 and the standard deviation was 76. In what subject was there greater variability?

Solution

English

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{80}{78} \times 100 = 102.56$$

Economics

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{73}{76} \times 100$$

$$= 97.6$$

The variabiity in English is more because C.V. in that case is more.

Illustration No.28

Calculate Mean, standard Deviation and coefficient of variation.

C.I.	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	4	8	8	16	12	6	4

Solution

$A = 45 \quad i = 10$

x	f	M.V.	dx	dx′	fdx′	fdx'^2
10-20	4	15	-30	-3	-12	36
20-30	8	25	-20	-2	-16	32
30-40	8	35	-10	-1	-8	8
40-50	16	45	0	0	0	0
50-60	12	55	+10	+1	+12	12
60-70	6	65	+20	+2	+12	24
70-80	4	75	+30	+3	+12	36
	Σf = 58				Σfdx′ = 0	$\Sigma fdx'^2$ = 148

$$\text{Mean} = A + \frac{\Sigma fdx'}{\Sigma f} \times i$$

$$= 45 + \frac{0}{58} \times 10 = 45$$

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx'^2}{\Sigma f} - \left(\frac{\Sigma fdx'}{\Sigma f}\right)^2} \times i$$

$$\text{S.D.} = \sqrt{\frac{148}{58} - \left(\frac{0}{58}\right)^2} \times 10$$

$$= 15.9$$

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{15.9}{45} \times 100 = 35.33$$

Illustration No.29

Obtain the standard deviation of the following frequency distribution of the population of 104 villages in a taluk. Calculate the co-efficient of variation also.

Population	No. of villages
0 - 200	10
200 - 400	28
400-1000	42
1000-2000	18
2000-5000	6

Solution

$A = 700 \qquad i = 200$

x	M.V. x	dx	dx′	f	fdx′	fdx'^2 $(fdx' \times dx')$
0- 200	100	-600	-3	10	-30	90
200- 400	300	-400	-2	28	-56	112
400-1000	700	0	0	42	0	0
1000-2000	1500	+800	+4	18	+72	288
2000-5000	3500	+2800	+14	6	+84	1176
				$\Sigma f =$ 104	$\Sigma fdx'$ $= 70$	$\Sigma fdx'^2$ $= 1666$

$$\text{Mean} = A + \frac{\Sigma fdx'}{\Sigma f} \times i$$

$$= 700 + \frac{70}{104} \times 200$$

$$= 834.67$$

$$\text{S.D.} = \sqrt{\frac{\Sigma fdx'^2}{\Sigma f} - \left(\frac{\Sigma fdx'}{\Sigma f}\right)^2} \times i$$

$$= \sqrt{\frac{1666}{104} - \left(\frac{70}{104}\right)^2} \times 200$$

$$= \sqrt{16.01 - 0.45} \times 200$$

$$= \sqrt{15.56} \times 200$$

$$= 3.94 \times 200$$

$$= 788$$

$$\text{C.V.} = \frac{788}{834.67} \times 100$$

$$= \frac{78800}{834.67}$$

$$= 94.408$$

LORENZ CURVE

American economist and statistician Prof. Max O Lorenz used a cumulative frequency graph to study the dispersion of the distribution of wealth and income. Since then, this graphical representation of a special type of frequency distribution is effectively used to study the distribution of profits, wages, output etc. This curve is most commonly used in studying inequality of income and wealth.

The following steps are taken to draw this curve:-

1. We first calculate the cumulative totals of the values of the variables of the distribution and then these cumulative totals are converted into percentages of the sum of the values of the variable.

2. Secondly, we calculate the cumulative frequencies of the distribution and then these cumulative frequencies are converted into percentages of the sum of the frequencies.

3. The percentage of the cumulative frequencies are represented on the X-axis and the percentage of the cumulative totals of the values of the variable are represented on the Y-axis.

4. The origin of the co-ordinate axis is taken as 0 on the X-axis and the scale marked on it from 0 to 100. The same origin is taken as 0 on the Y-axis and the scale marked on it from 0 to 100.

5. The points representing (0, 0) and (100, 100) are joined with the help of a straight line. This line is known as *"line of Equal Distribution"*. Any point on this line shows the same percentage on X-axis as on Y-axis.

6. Plot the various points corresponding to the values of the variables x and y and then join them. The graph of the distribution so obtained, unless it is exactly equal, will always be a curve below the line of equal distribution.

It may be noted that if two curves of two distributions are shown on the same Lorenz diagram, the curve that is farthest from the line of equal distribution willl show greater inequality of the distribution than the other distribution.

Illustration No.30

From the following figures, prepare a Lorenz Curve and find out which group has got the greater inequality:

Profit in Rs.	100	200	300	400	500	600	900	1000
Group A- No. of firms	18	24	27	30	21	15	9	6
Group B- No. of firms	21	33	36	42	24	18	14	12

Solution

Profit			Group A			Group B		
Rs.	Cumulative Profit	% of c.f.	No. of firms	c.f.	% of cf	No. of firms	c.f. of	% of c.f.
100	100	2.5	18	18	12	21	21	10.5
200	300	7.5	24	42	28	33	54	27.0
300	600	15.0	27	69	46	36	90	45.0
400	1000	25.0	30	99	66	42	132	66.0
500	1500	37.5	21	120	80	24	156	78.0
600	2100	52.5	15	135	90	18	174	87.0
900	3000	75.0	9	144	96	14	188	94.0
1000	4000	100.0	6	150	100	12	200	100.0

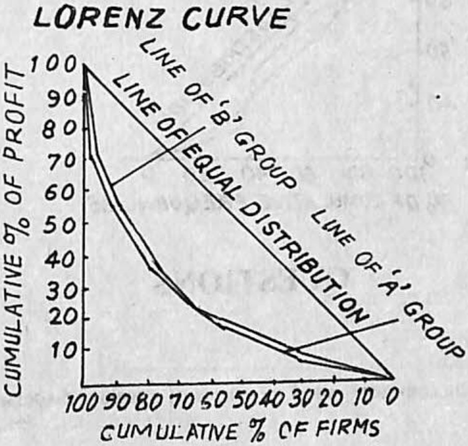


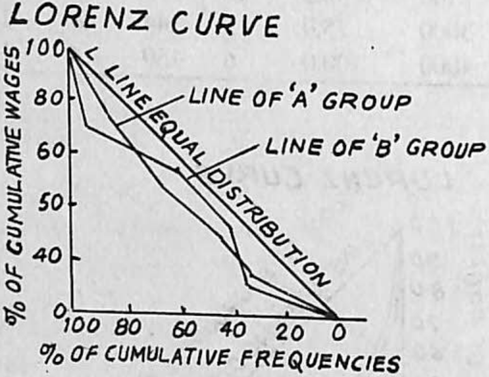
Illustration No.31

Given below are the figures of the wages of the workers in two groups of factories A and B

Inocme in Rs	100-200	2 00-300	300-400	400-500	500-600	600-700
No. of workers Factory A	130	150	160	200	190	170
No. of workers Factory B	300	320	120	460	600	200

Solution

Wages In Rs.				Factory A			Factory B		
C.I	. M.V.	c.f.	% of c.f.	No. of workers	c.f.	% of. c.f.	No. of workers	c.f.	% of c.f.
100-200	150	150	6.25	130	130	13	300	300	15
200-300	250	400	16.67	150	280	28	320	620	31
300-400	350	750	31.25	160	440	44	120	740	37
400-500	450	1200	50.00	200	640	64	460	1200	60
500-600	550	1750	73.00	190	830	83	600	1800	90
600-700	650	2400	100.00	170	1000	100	200	2000	100



QUESTIONS

1. Name four commonly used measures of absolute dispersion. (Hint: Range, Q.D., M.D., S.D.)

2. Name the most commonly used measure of relative dispersion. Give the formula for calculating it.
(Hint: Coefficient of variation).

3. Find the standard deviation of the height of 100 students.

Height in inches	Frequency
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

4. Find the standard deviation and mode from the following data:-

Wages	No. of workers
upto 10	12
20	30
30	65
40	107
50	157
60	202
70	222
80	230

5. Calculate the mean deviation and co-efficient of mean deviation from the following data:-

Marks	No. of students
0-10	4
0-20	6
20-30	8
30-40	10
40-50	7
50-60	6
60-70	5
70-80	3
80-90	1

6. Write notes on the following:

a) Standard Deviation

b) Lorenz Curve

7. Calculate mean and standard deviation from the following distribution:-

Item	10-20	20-30	30-40	40-50	50-60	60-70	70-80
frequency	4	8	8	16	12	6	4

8. Calculate the mean deviation from the following data:-

C.I.	0-20	20-40	40-60	60-80	80-100	100-200
Frequency	5	50	84	32	10	6

9. Define Lorenz Curve.

In the following table is the number of companies belonging to two areas A and B acccording to the amounts of profits earned by them. Draw their Lorenz Curve:-

Profits earned in Rupees	Area A	Area B
6000	6	2
25000	11	38
60000	13	52
84000	14	28
105000	15	38
150000	17	26
170000	10	12
400000	14	4

10. Calculate the Arithmetic Mean and Standard Deviation from:-

C.I.	0-10	10-20	20-30	30-40	40-50
Frequency	8	13	16	8	5

11. Obtain the standard deviation of the following frequency distribution of the population of 104 villages in a taluk.

Population	No. of villages
0 - 200	10
200 - 400	28
400 - 1000	42
1000-2000	18
2000-5000	6
	104

12. (a) What are the different measures of 'income inequality'?

Evalute critically their relative strength and weakness.

(b) 9,400 Indian households are classified according to their after-tax income as follows:-

After tax income:	0-1000	1000-5000	5000-10000	10000-20000	20000-40000
No. of households:	1348	4210	1892	1460	490

Draw the Lorenz Curve for these households. What conclusions can you draw from the above data about the inequality of income distribution?

(Hint: The most appropriate device will be Lorenz Curve. Other devices are

Range, Q.D., M.D., S.D.Range is difference of smallest and largest.Q.D. is half of the difference between Q.3 and Q.1.

M.D. is the average of deviations from the mean ignoring plus and minus signs. The greatest shortcoming is to ignore the plus and minus signs. S.D. is considered good method of dispersion).

13. Compute the standard deviation of the amount of money held for the following individuals.

Individual	1	2	3	4	5	6	7	8	9	10
Money (Rs)	114	108	100	98	101	109	117	119	121	126

14. Outline in brief the necessity of having a measure of dispersion when summarising a set of data:

(Hint: an average gives only a single representative figure about a given set of data. As such, one does not know anything about the composition of a series i.e. how different items of a series are dispersed vis-a-vis of a series i.e. how different items of a series are dispersed vis-a-vis average. Thus, with the help of disperation, one gets an idea about the extent of variation of different items of a series in relation to its average.)

15. Calculate the arithemetic mean and standard deviation from the following series-

C.I.	5-15	15-25	25-35	35-45	45-55
Frequency	8	12	15	2	6

16. Write short notes on:-

(a) Lorenz Curve

(b) Quartiles

17. (a) The means and standard deviations of the incomes of teachers in two schools A and B are given below;-

School	Mean income(Rs.)	Standard deviation
A	800	120
B	900	135

Compare the variability of incomes of teachers in two schools.

(B) Calculate the variance of the numbers 2, 7 and 9.

18. Find out the mean and standard deviation from the following distribution:-

0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
2	5	7	13	21	16	8	3

19. (a) What do you understand by mean deviation?
(b) In what way is mean deviation a better measure of disperasion than range?
(c) Calculate mean deviation for the following data on the monthly income (in rupees) of five households.

825 635 792 836 750

(Hint: (b) Mean deviation is better than range because range is concerned with two extreme values where as M.D. is based on all observations. Range is silent about what happens to values between two extremes. Range also fluctuates much from sample to sample).

20. Find out mean and standard deviation from the following distribution:-

C.I.	70-80	60-70	50-60	40-50	30-40	20-30
Frequency	7	11	22	0	15	5

(Hint: Regroup the order as 20-30, 30-40 and 70-80)

21. What is standard deviation? Find out the standard deviation. With what symbol is it shown?

C.I.	5-15	15-25	25-35	35-45	45-55
Frequency	8	12	15	9	6

(Hint; S.D. is a measure of dispersion. It is best measure of dispersion because it is free from these defects of which other measure of dispersion suffer. S.D. is square root of the mean of the sum of the squares from the mean. It is also known as root mean square deviation).

22. Find out the standard deviation of the folowing distribution;

Age in Years	50-55	45-50	40-45	35-40	30-35	25-30
No. of labour	22	29	31	47	51	70

23. Find Out the mean and Standard deviation from the following:-

0-4	4-8	8-12	12-16
4	8	2	1

6

Measures of Correlation (Two variables and Ungrouped Data)

Measures of Central tendency describes the nature of distribution relating to a single variable. An economist may be interested in studying relationship between two and more variables e.g. income and consumption are related variables; similarly price and demand are related variables, productivity and wage rates depend upon each other. There can be three types of relationship e.g.(i) an increase/decrease in the values of one variable may result in the increase/ decrease in the values of the other variable, such relation is called positive (ii) the increase or decrease in the values of one variable may result in the decrease or increase in the values of the other variable, such relationship is called inverse or negative (iii) the increase or decrease in the values of one variable may not effect the values of the other variable. This effect or interrelationship is called non-relationship. The extent or degree of relationship can be measured with the help of a statistical tool which is called correlation.

Correlation may be defined as under:

"The statistical tool with the help of which relationship between two inter-related variables is measured is called correlation".

Types of Correlation:

i) Positive and Negative Correlation: When both the variables change in the same direction, that is when both increase or decrease, the relationship between the two variables is called positive or direct. But when the change is in opposite directions, that is one is increasing and the other is decreasing, the correlation is negative or inverse. Examples of positive correlation can be increase in heat affecting temperature, increase in advertisement affecting sale. Increase in height, affecting weight, increase in rainfall increasing production of rice etc.

Examples of negative correlation can be rise in prices causing fall in demand. Increase in the number of television sets, decreases the number of cinema goers. Day temperature rising, sale of woollen garments reducing.

ii) Perfect Correlation : Perfect correlation occurs where change in two variables is exactly proportional. If equal proportional changes are in the same direction, there is perfect positive correlation between the two variabiles. If equal proportional changes are in the reverse direction, there is perfect negative correlation. The value of correlation coefficient will be ±1 in these cases.

iii) No Correlation : When there exists no inter-dependence between the two variables or no relationship is found between the change in them, there is no correlation or absence of correlation, described as zero(0).

iv) Limited Degree of Correlation : In social sciences, the variables may be correlated, but an increase in one variable need not always be accompained by a corresponding or equal increase (or decrease) in the other variable.

Correlation is said to be limited positive when there are unequal changes in two variables in the same direction and correlation is limited negative when there are unequal changes in the reverse direction.

Methods of finding out Correlation

There are various methods of calculating the correlation between two variables out of which the most commonly used are:

i) Scatter Diagram

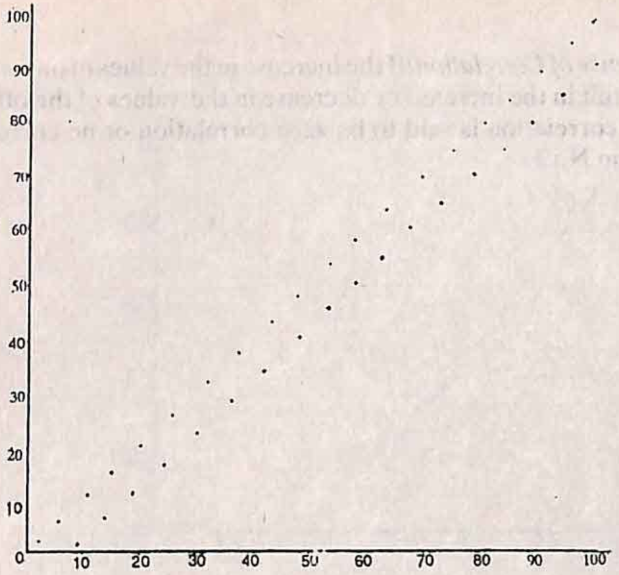
ii) Karl Pearson's Co-efficient of Correlation: and

iii) Spearmans' Rank Correlation Co-efficient.

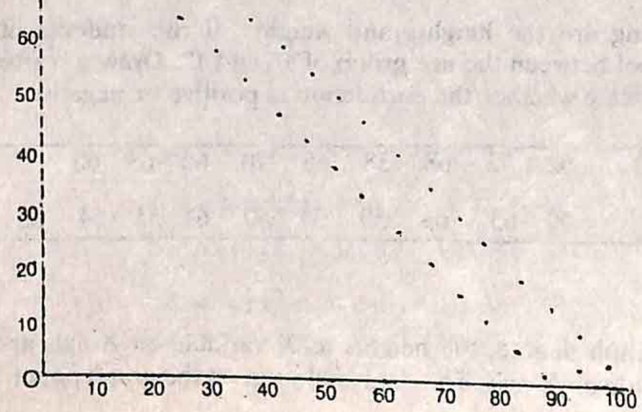
SCATTER DIAGRAM: The simplest device for studying the inter-relationship between two variables is know as scatter diagram. In this method, we plot on the graph paper the corresponding pairs of values of the two variables, one may be taken as independent variable and the other dependent variable. The values of the independent variable are taken on the horizontal axis and the values of the dependent variable are taken on the vertical axis. The values are plotted on the graph paper in the form of dots. This graphical representation of the values of the variables is known as *scatter diagram* or *dot diagram*. By observing the pattern of dots we can know the presence or absense of correlation and the type of correlation.

(i) Direct or Positive Correlation: If the values of one variable are associated with the increase of values on the other variable, the correlation is said to be direct or positive.

See diagram No.1. The dots in the scatter diagram will cluster around a straight line going upwards from left to right.



ii) Indirect or Negative Correlation: If the increase in the values of one variable results in the decrease in the values of the other variable or vice-versa such a correlation is said to be indirect or negative. See diagram No.2



iii) Absense of Correlation: If the increase in the values of one variable does not result in the increase or decrease in the values of the other variable, such a correlation is said to be zero correlation or no correlation. See Diagram No.3.

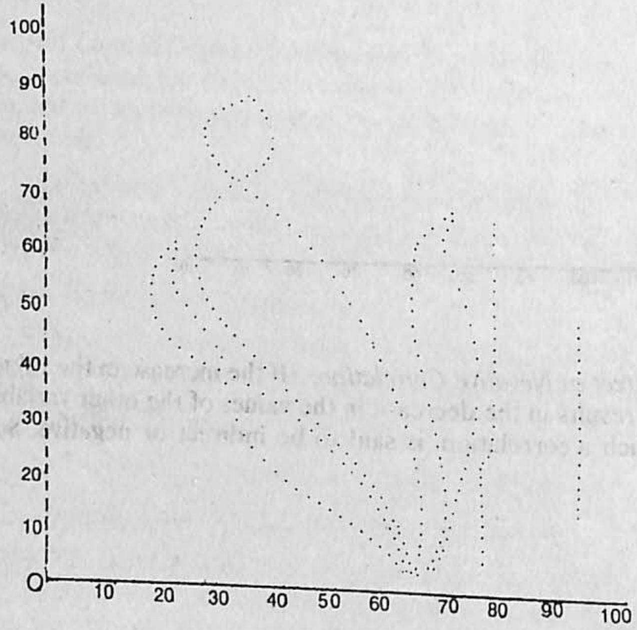


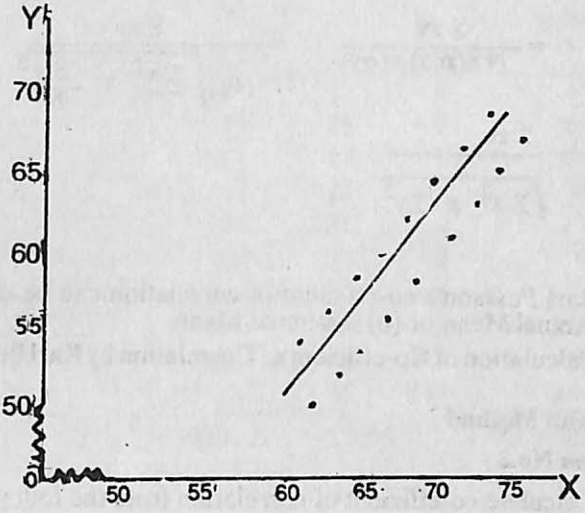
Illustration No.1

Following are the heights and weights of the students of Cambridge school between the age group of 16 and 17. Draw a scatter diagram and indicate whether the correlation is positive or negative.

Height:(Inches)	62	72	68	58	65	70	66	63	60	72
Weight(Kg):	50	65	63	50	54	60	61	55	54	65

Solution

In the graph denote, the heights as X variable on X-axis and weight as Y variable on Y-axis. The data is plotted on the graph paper.



Correlation between the two variables is positive, since most of the dots cluster around the line moving upwards.

KARL PEARSON'S CORRELATION CO-EFFICIENT

A famous statistician, Karl Pearson used a mathematical technique to compute the degree of correlation between two variables. It is represented by the symbol 'r'. The co-efficient of correlation (r) of two variables is obtained by dividing the sum of the products of corresponding deviations of the various items of two series from their respective means by the product of their standard deviations and the number of pairs of observations.

Symbolically,

$$r = \frac{\Sigma xy}{N \times (\sigma x) \times (\sigma y)}$$

Here, $x = (x - \bar{x})$; $y = (y - \bar{y})$

σx = Standard deviation of x series.

σy = Standard deviation of y series.

N = Number of pairs of observations.

r = co-efficient of Correlation.

The above formula can be re-written as under:

$$r = \frac{\Sigma xy}{N \times (\sigma x) \times (\sigma y)} = \frac{\Sigma xy}{N x \sqrt{\frac{\Sigma x^2}{N} x \frac{\Sigma y^2}{N}}}$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Karl Pearson's co-efficient of correlation can be calculated by using (a) Actual Mean or (b) Assumed Mean.

Calculation of Co-efficient of Correlation by Karl Pearson's Formula:

Actual Mean Method

Illustration No.2

Calculate co-efficient of correlation from the following data:

X	15	18	21	24	27	30	36	39	42	48
Y	25	25	27	27	31	33	35	41	41	45

Solution

Steps:

1. Calculate arthmetic mean of X and Y series.
2. Find out the deviations of X and Y series from the means of X and Y and denote them by "x" and "y".
3. Square these deviations and obtain the totals i.e. Σx^2 and Σy^2.
4. Multiply the deviations as obtained in step 2 above to get xy.
5. Add up the column xy to obtain Σxy.
6. Apply the Formula:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

X	$(X-\overline{X})=X$	X^2	Y	$(Y-\overline{Y})=Y$	Y^2	XY
1	2	3	4	5	6	(2 x 5)
15	−15	225	25	−8	64	120
18	−12	144	25	−8	64	96
21	−9	81	27	−6	36	54
24	−6	36	27	−6	36	36
27	−3	9	31	−2	4	6
30	0	0	33	0	0	0
36	+6	36	35	+2	4	12
39	+9	81	41	+8	64	72
42	+12	144	41	+8	64	96
48	+18	324	45	+12	144	216
$\Sigma X = 300$		$\Sigma x^2 = 1080$	$\Sigma Y = 330$		$\Sigma Y^2 = 480$	$\Sigma xy = 708$

$$\text{Mean } (\overline{X}) = \frac{300}{10} = 30$$

$$\text{Mean } (\overline{Y}) = \frac{330}{10} = 33$$

Applying formula we get:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$r = \frac{708}{\sqrt{1080 \times 480}} = \frac{708}{\sqrt{518400}}$$

$$= \frac{708}{720} = 0.98$$

Here there is high degree of positive correlation.

Illustration No. 3

Calculate the co-efficient of correlation between birth rate and death rate from the data given on the next page.

Year	Birth Rate	Death Rate
1931	24	15
1941	26	20
1951	32	22
1961	33	24
1971	35	27
1981	30	24

Solution

1	2	3	4	5	6	
X (Birth rate)	Deviation from 30 X	X^2	Y (Death rate)	Deviation from 22 Y	Y^2	xy (2 x 5)
24	− 6	36	15	− 7	49	42
26	− 4	16	20	− 2	4	8
32	+ 2	4	22	0	0	0
33	+ 3	9	24	+ 2	4	6
35	+ 5	25	27	+ 5	25	25
30	0	0	24	+ 2	4	0
$\Sigma X =$ 180	$\Sigma x = 0$	Σx^2 $= 90$	$\Sigma Y =$ 132	$\Sigma y = 0$	Σy^2 $= 86$	Σxy $= 81$

$$\text{Mean (X)} = \frac{180}{6} = 30 \qquad \text{Mean (Y)} = \frac{132}{6} = 22$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \text{ x } \Sigma y2}}$$

$$= \frac{81}{\sqrt{90 \text{ x } 86}}$$

$$= \frac{81}{87.9} = 0.91$$

Here the degree of correlation is high and positive.

Assumed Mean Method

Illustration No.4

Calculate Karl Pearsons Co-efficient of Correlation between the values ot x and y given below

X	78	89	97	69	59	79	68	61
Y	125	137	156	112	107	136	123	108

Solution

X	Deviation from Assumed Mean 69 x	x^2	Y	Deviation from Assumed Mean 112 y'	y^2	xy
78	+9	81	125	+13	169	+117
89	+20	400	137	+25	625	+500
97	+28	784	156	+44	1936	+1232
69	0	0	112	0	0	0
59	– 10	100	107	– 5	25	+50
79	+10	100	136	+24	576	+240
68	–1	1	123	+11	121	–11
61	–8	64	108	–4	16	+32
n = 8	$\Sigma x =$ +48	$\Sigma x^2 =$ 1530	n = 8	$\Sigma y =$ +108	$\Sigma y^2 =$ 3468	$\Sigma xy =$ 2160

$$r = \frac{\Sigma xy - \frac{1}{n}\,\Sigma x\,\Sigma y}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \times \sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

$$= \frac{2160 - \frac{1}{8}\,(48 \times 108)}{\sqrt{1530 - \frac{(48)^2}{8}}\,\sqrt{3468 - \frac{(108)^2}{8}}}$$

$$= \frac{2160 - 648}{\sqrt{1530 - 288}\,\sqrt{3468 - 1458}}$$

$$= \frac{1512}{\sqrt{1242} \times \sqrt{2010}} = \frac{1512}{1576} \text{ (app)}$$

$= 0.96$ (app.).

There is a high degree of positive corelation.

Illustration No.5

Calculate Karl Pearson's Co-efficient of Correlation between values of x and y given below:-

x	5	15	12	20	18	25	16	30	32	35
y	1	4	3	6	5	7	4	8	9	8

Solution

A=20				A=5		
x	dx	dx^2	y	dy	dy^2	(dx) (dy)
5	–15	225	1	–4	16	60
15	–5	25	4	–1	1	5
12	–8	64	3	–2	4	16
20	0	0	6	+1	1	0
18	–2	4	5	0	0	0
25	+5	25	7	+2	4	10
16	–4	16	4	–1	1	4
30	+10	100	8	+3	9	30
32	+12	144	9	+4	16	48
35	+15	225	8	+3	9	45
n=10	Σ dx =8	Σdx^2 =828		Σdy = 5	Σdy^2 =61	Σdxdy =218

$$r = \frac{\Sigma dx \times dy - \dfrac{(\Sigma dx)\ (\Sigma dy)}{n}}{\sqrt{\left[\Sigma dx^2 - \dfrac{(\Sigma dx)^2}{n}\right] \quad \left[\Sigma dy^2 - \dfrac{(\Sigma dy)^2}{n}\right]}}$$

$$= \frac{218 - \dfrac{8 \times 5}{10}}{\sqrt{\left[828 - \dfrac{(8)^2}{10}\right]\left[61 - \dfrac{(5)^2}{10}\right]}}$$

$$= \frac{214}{\sqrt{[828 - 6.4]\ [61 - 2.5]}}$$

$$= \frac{214}{\sqrt{821.6 \times 58.5}} = 0.98$$

There is high degree of correlation.

Illustration No. 6

Calculate the coefficient of correlation between the values of the following X and Y.

X	78	89	96	69	59	79	68	61
Y	125	137	156	112	107	136	123	108

Solution

A = 75 A= 125

X	dx	dx^2	Y	dy	dy^2	dx x dy
78	+3	9	125	0	0	0
89	+14	196	137	+12	144	168
96	+21	441	156	+31	961	651
69	−6	36	112	−13	169	78
59	−16	256	107	−18	324	288
79	4	16	136	+11	121	44
68	−7	49	123	−2	4	14
61	−14	196	108	−17	289	238
$\Sigma x = 599$	$\Sigma dx = -1$	$\Sigma dx^2 =$ 1199	$\Sigma y =$ 1,004	$\Sigma dy = 4$	Σdy^2 = 2012	$\Sigma dx \times dy$ = 1481

Apply formula we get:

$$r = \frac{\Sigma dx \times dy - \dfrac{(\Sigma dx) \times (\Sigma dy)}{N}}{\sqrt{\left[\Sigma dx^2 - \dfrac{(\Sigma dx)^2}{N}\right]\left[\Sigma dy^2 - \dfrac{(\Sigma dy)^2}{N}\right]}}$$

$$r = \frac{1481 - \dfrac{(-1)\ (4)}{8}}{\sqrt{\left[1199 - \dfrac{(-1)^2}{8}\right]\left[2012 - \dfrac{(4)^2}{8}\right]}}$$

$$= \frac{1481.5}{\sqrt{(1198.875)\ (2010)}}$$

$$= \frac{1481.5}{1552.333}$$

0.95

There is high degree of positive correlation between X and Y.

Illustration No. 7

Calculate co-efficient of correlation between x and .

X	10	12	8	15	20	25	40
Y	15	10	6	25	16	12	8

Solution

A = 20 A = 16

1. x	2. dx	3. dx^2	4. Y	5. dy	6. dy^2	7. dx x dy
10	–10	100	15	– 1	1	10
12	– 8	64	10	– 6	36	48
8	– 12	144	6	– 10	100	120
15	– 5	25	25	+9	81	- 45
20	0	0	16	0	0	0
25	5	25	12	– 4	16	– 20
40	+20	400	8	– 8	64	– 160
N = 7	Σdx = –10	Σdx^2 = 758		Σdy = -20	Σdy^2 = 298	Σdx x dy = –47

Applying formula,we get

$$r = \frac{\Sigma dx \times dy - \dfrac{\Sigma dx \times \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \dfrac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \dfrac{(\Sigma dy)^2}{N}}}$$

$$= \frac{-47 - \dfrac{-10 \times -20}{7}}{\sqrt{758 - \dfrac{(-10)^2}{7}} \times \sqrt{298 - \dfrac{(-20)^2}{7}}}$$

$$= \frac{-47 - 28.57}{\sqrt{758 - 14.285}\sqrt{298 - 57.142}}$$

$$= \frac{-75.57}{\sqrt{743.715} \times \sqrt{240.8485}}$$

$$= \frac{-75.57}{423.218} = -0.178$$

Here there is low degree of negative correlation between x and y values.

Illustration No.8

Calculate the coefficient of correlation that exist between price and quantity and comment on it.

Price	Quantity
10	5
10	6
11	4
12	3
12	3

Solution

A = 11 A = 4

x	dx	dx^2	y	dy	dy^2	dx dy
10	–1	1	5	+1	1	–1
10	–1	1	6	+2	4	–2
11	0	0	4	0	0	0
12	+1	1	3	–1	1	–1
12	+1	1	3	–1	1	–1
N = 5	$\Sigma dx = 0$	$\Sigma dx^2 = 4$		$\Sigma dy = 1$	$\Sigma dy^2 = 7$	$\Sigma dxdy = -5$

By applying formula we get,

$$r = \frac{\Sigma dxdy - \frac{(\Sigma dx)\ (\Sigma dy)}{N}}{\sqrt{\left[\ \Sigma dx^2 - \frac{(\Sigma dx)^2}{N}\right]\left[\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}\right]}}$$

$$= \frac{-5 - \frac{0 \times 1}{5}}{\sqrt{\left[4 - \frac{(0)^2}{5}\right]\left[7 - \frac{(1)^2}{5}\right]}}$$

$$= \frac{-5}{\sqrt{4 \times 6.8}}$$

$$= \frac{-5}{\sqrt{27.2}}$$

$$= \frac{-5}{5.215}$$

$$= -0.958$$

There is high degree of negative correlation between price and quantity.

SPEARMANS' RANK CORRELATION

British psychologist Prof. Charles Spearman has conceived another method of finding co-efficient of correlation. The co-efficient is expressed as R and is based on the ranking of the various values of the two variables. This method is specially useful where certain factors can not be precisely measured in quantitative terms as in case of intelligence, leadership quality, sense of cooperation and degree of honesty etc.

After assigning ranks to the various items, the difference of the corresponding rank values are calculated and then correlation is found out with the help of following formula:

$$R = -1 - \frac{6 \Sigma D^2}{N(N^2 - 1)}$$

Illustration No.9

Find rank correlation co-efficient from the following data:-

X	8	36	98	25	75	82	92	62	65	35
Y	84	51	91	60	68	62	86	58	35	49

Solution

Steps

1. Determine the ranks of X and Y.

2. Take deviations of the ranks of X from the ranks of Y and denote these as D.

3. Square the deviations and add them to get ΣD^2.

4. Apply formula

$$R = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

X	RANK	Y	RANK	D	D^2
8	10	34	3	7	49
36	7	51	8	-1	1
98	1	91	1	0	0
25	9	60	6	3	9
75	4	68	4	0	0
82	3	62	5	-2	4
92	2	86	2	0	0
62	6	58	7	-1	1
65	5	35	10	-5	25
35	8	49	9	-1	1
					$\Sigma D^2 = 90$

$$R = 1 - \frac{6 \Sigma D^2}{N(N^2 - 1)}$$

$$= 1 - \frac{6 \times 90}{10(100 - 1)}$$

$$= 1 - \frac{540}{990} = \frac{5}{11} = 0.45$$

When two items have same values, the corresponding ranks will be divided between them as explained in the following examples:

Illustration No.10

X	50	50	55	60	65	65	65	60	60	50
Y	13	15	16	18	18	17	17	16	15	15

X	Rank of X	Y	Rank of Y	D	D2
50	9	13	10	1.0	1.00
50	9	15	8	1.0	1.00
55	7	16	5.5	1.5	2.25
60	5	18	1.5	3.5	12.25
65	2	18	1.5	0.5	0.25
65	2	17	3.5	–1.5	2.25
65	2	17	3.5	–1.5	2.25
60	5	16	5.5	– 0.5	0.25
60	5	15	8	– 3.0	9.00
50	9	15	8	+ 1.0	1.00
				$\Sigma D = 0$	$\Sigma D^2 =$ 31.50

$$65 = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

$$60 = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

$$50 = \frac{8+9+10}{3} = \frac{27}{3} = 9$$

$$18 = \frac{1+2}{2} = \frac{3}{2} = 1.5$$

$$17 = \frac{3+4}{2} = \frac{7}{2} = 3.5$$

$$16 = \frac{5+6}{2} = \frac{11}{2} = 5.5$$

$$15 = \frac{7+8+9}{3} = \frac{24}{3} = 8$$

Apply formula

$$R = 1 = \frac{6\Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 31.50}{10\ (100-1)}$$

$$= 1 - \frac{189}{990} = \frac{89}{110}$$

$$= 0.8$$

Illustration No. 11

Calculate the rank order correlation between price and supply:

Price:	4	6	8	10	12	14	16	18
Supply:	10	15	20	25	30	35	40	45

Solution

X Price	Rank of X	Supply y	Rank of Y	D	D2
4	8	10	8	0	0
6	7	15	7	0	0
8	6	20	6	0	0
10	5	25	5	0	0
12	4	30	4	0	0
14	3	35	3	0	0
16	2	40	2	0	0
18	1	45	1	0	0
N=8				$\Sigma D = 0$	$\Sigma D^2 = 0$

Apply formula

$$R = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 0}{8(64-1)}$$

$$= 1$$

X	Rank of X	Y	Rank of Y	D	D2
50	9	13	10	1.0	1.00
50	9	15	8	1.0	1.00
55	7	16	5.5	1.5	2.25
60	5	18	1.5	3.5	12.25
65	2	18	1.5	0.5	0.25
65	2	17	3.5	–1.5	2.25
65	2	17	3.5	–1.5	2.25
60	5	16	5.5	– 0.5	0.25
60	5	15	8	– 3.0	9.00
50	9	15	8	+ 1.0	1.00
				$\Sigma D = 0$	$\Sigma D^2 =$ 31.50

$$65 = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

$$60 = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

$$50 = \frac{8+9+10}{3} = \frac{27}{3} = 9$$

$$18 = \frac{1+2}{2} = \frac{3}{2} = 1.5$$

$$17 = \frac{3+4}{2} = \frac{7}{2} = 3.5$$

$$16 = \frac{5+6}{2} = \frac{11}{2} = 5.5$$

$$15 = \frac{7+8+9}{3} = \frac{24}{3} = 8$$

Apply formula

$$R = 1 = \frac{6\Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 31.50}{10\ (100-1)}$$

$$= 1 - \frac{189}{990} = \frac{89}{110}$$

$$= 0.8$$

Illustration No. 11

Calculate the rank order correlation between price and supply:

Price:	4	6	8	10	12	14	16	18
Supply:	10	15	20	25	30	35	40	45

Solution

X Price	Rank of X	Supply y	Rank of Y	D	D2
4	8	10	8	0	0
6	7	15	7	0	0
8	6	20	6	0	0
10	5	25	5	0	0
12	4	30	4	0	0
14	3	35	3	0	0
16	2	40	2	0	0
18	1	45	1	0	0
N = 8				$\Sigma D = 0$	$\Sigma D^2 = 0$

Apply formula

$$R = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$= 1 - \frac{6 \times 0}{8(64-1)}$$

$$= 1$$

QUESTIONS

1. Calculate the co-efficient of correlation between the population of two cities:

Year	City A (in lakhs)	City B (in lakhs)
1921	340	260
1931	300	280
1941	360	330
1951	370	310
1961	380	300
1971	390	320
1981	380	300

(Hint:- Treat city A as X variable and city B as Y variable, Ignore the years. Mean of X=360, Y=300. Take deviations from mean, square them, multiply x and y i.e. deviation and find out Σxy. Apply formula:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

Ans = 0.6

2. What is meant by correlation? Calculate co-efficient of correlation between the values of the following variables x and y:-

X	20	24	16	36	44	48	70
Y	30	20	12	52	34	26	20

(Hint: By applying formula:

$$r = \frac{\Sigma dx \times dy - \frac{(\Sigma dx)(\Sigma dy)}{N}}{\sqrt{\left[\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}\right]\left[\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}\right]}} = 0.093$$

3. Calculate the co-efficient of correlation between the birth and death rate from the following data:-

Year	Birth rate	Death rate
1931	24	15
1941	26	20
1951	32	22
1961	33	24
1971	35	27
1981	30	24

(Hint: See illustration No.3) Ans = .91

4. Give three examples of perfect correlation. Calculate the co-efficient of Rank order correlation from the following:

X	48	33	40	9	16	16	65	25	15	57
Y	13	13	22	6	14	4	20	9	6	15

Ans = 0.745

(Hint: Examples of perfect correlation are-

i) Income and expenditure

ii) Height and Weights

iii) Ages of wives and husbands)

5. Write short notes on:-

a) Co-efficient of correlation

b) Positive and negative correlation

c) Rank correlation

6. Calculate the co-efficient of rank correlation from the following data:

X	10	12	8	15	20	25	40
Y	15	10	6	25	16	12	18

Ans = 0.57

7. The following table gives the height and weight of 10 students in a class. Draw a scatter diagram and interpret whether the correlation is positive or negative:-

Height(cm)	180	150	158	165	175	163	145	195	180	155
Weight(Kgs)	65	54	55	61	60	54	50	63	65	50

(Hint: Positive correlation-degree does not seem to be very high)

8. Calculate co-efficient of correlation between X and Y:

X	10	12	8	15	20	25	40
Y	15	10	6	25	16	12	8

Ans = 0.18

9. The data on price and quantity purchased relating to commodity for 5 months are given below: Calculate co-efficient of correlation between price and quantity.

Months:	1	2	3	4	5
Price (Rs)	10	10	11	12	12
Quantity (Kg)	5	6	4	3	3

(Hint: Ignore months-Assume Mean and solve the question) Ans = 0.96

10. Calculate the co-efficient of Rank correlation between price and supply from the following data:

Price	4	6	8	10	12	14	16	18
Supply	10	15	20	25	30	35	40	45

Ans = 1

11. Write the utility of co-efficient of rank correlation.

(Hint: This method is simple to understand and easily applicable. This Method should not be used with frequency distributions and where number exceeds 30).

12. Make a scatter diagram from the following and interpret the results.

X	2	3	5	6	8	9
Y	6	5	7	8	12	11

13. What do you mean by scatter diagram. How is it used to determine correlation?

14. What are the various methods to determine correlation?

15. Distinguish between positive and negative correlation.

7

Index Numbers

"Index numbers are economic barometers. They are used to measure the pulse of the economy. They are used as indicators of inflationary and deflationary tendencies in the economy. When the economists want to make a statement about the rise or fall in prices of commodities, rise or fall in the cost of living, they make use of a statistical device called *Index Numbers*. Generally, we read in the newspapers that the dearness allowance of Government Employees is increased, when consumers price index goes up. Similarly, we come across index numbers of agriculatural and industrial production, sales, exports, prices,wages etc. They are useful to assess the development of an economy. They facilitate a comparision and analysis of unlike series such as employment, prices and production by establishing a base period and converting subsequent numbers into percents and indices.

Various economists have given various definitions for Index Numbers which are as under:

Secrist: "Index numbers are a series of numbers by which changes in the magnitude of a phenomenon are measured from time to time or from place to place."

Bowley says, "A series of index numbers is a series which reflects trend and fluctutations in the movements of same quantity to which it is related."

Croxton and Cowden, "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

Thus, we can say that an index number is a statistical tool lto measure the relative change in the magnitude of group of related variables, from time to time or from place to place.

Characteristics of Index Numbers

Following are some of the special characteristics of Index Numbers:

1. Index Numbers are numerical measures:Index numbers are expressed in terms of pecentages so as to show the extent of relative change.

Comparison becomes simple and easy. We generally do not use the percentage sign with index numbers.

2. *Index numbers are relative or comparative measurement of Group of Items:*Index numbers by their nature are comparative. They compare changes taking place overtime, or between place or like categories. The cost of living of the resident of Delhi in 1988 can be compared with the cost of living in 1978. If whole sale price index number of 1988 is 250 as compared to 100 of 1980, it means price level of 1988 has increased by 2.50 times.

3. *Index numbers are specialised averages:*According to M.M. Blair, "Index Numbers are specialised types of averages." Generally averages are used to compare only those series which are expressed in the same units, but device of index numbers helps us in comparing change in series which are in different units. Therefore, index numbers are called specialised type of averages.

4. *Index Numbers measure changes not capable of direct Measurement:*The technique of index numbers is utilised in measuring changes in magnitude which are not capable of direct measurement due to composite and complex character of the phenomenon. Changes in business activity in a country are not capable of direct measurement but it is possible to study relative changes in business activity by studying the variation in the values of some such factors which affect business activity and which are capable of direct measurement.

USES OF INDEX NUMBERS

Index numbers are guide post for the policy makers and the businessmen. They are invaluable for economists, as they are called Economic Barometer. They are helpful to feel the pulse of the economy. Following are some of their uses.

(a) Index Numbers guide the governments to take policy decisions such as to increase the dearness allowanace in the time before the situations leading to strikes or lock-outs arises.

(b) Index numbers help to measure the intelligence quotients, the changes in population in cities or towns, the changes in the adequacy of hospital facilities, the industrial or agriculatural production and so on.

(c) Index numbers help to measure value of money. On the basis of index numbers the value of money is found out. If the prices are more the value of money will be less and vice-versa. We come to know about the standard of living of the people.

(d) Index numbers enable us to measure real wages and standard of living. They tell us about the change in wages and standard of living. We can also make comparisons in the wages by taking different years into consideration.

(e) Index numbers are helpful in foreign trade. We can improve the balance of payments of a country by having the relevant index numbers.

(f) Index numbers help in providing us the idea about the rate of growth of the industry. The production in an economy is increasing or decreasing.

(g) Index numbers tells us about the changes in the value of money. On that basis the central bank of the country (Reserve Bank of India for India) can formulate proper monetary policy.

(h) Index numbers are very helpful in deflating (adjusting) national income on the basis of constant prices, to enable us to find out whether there is any change in the real income of the people.

(i) Index numbers help us to find out the trend of exports, imports, balance of payments, industrial production, prices, national income and variety of other phenomena. They are also useful in forecasting future trends. With the help of index numbers, a business executive is in a better position to take decisions about whether a new prodcut should be launched or whether there is any possibility of exploring new markets or whether the existing price policy needs any revision.

(j) Index numbers help in formulating the taxation policy. They tell us which taxes could be yielding better results.

(k) Index numbers helps a businessman to know about the increase or fall in sales etc.

(l) Index numbers are very helpful to politicians. Index numbers are the index of economic problems and the politicians are very much interested to solve economic problems. With the figures, supplied by index numbers, politicians can raise their voice in parliament and legislative assemblies.

(m) Index numbers measure the pulse of an economy and acts as barometers to find the ups and downs in general economic conditions of a country. Like barometers, which are used in physics so index numbers are used to measure the level of economic and business activity in the country. They can measure prices, output, foreign exchange, resources of the Reserve Bank of India etc.

Limitations of Index Numbers:

Lord Keynes has pointed out certain limitations of index numbers. The chief limitations are:

(i) Index numbers are different in different countries so international comparison is not possible.

(ii) It is very difficult to compare the prices at different times with the help of index numbers because the form of consumption changes with the change in time. People consume new goods in place of old goods.

(iii) Index numbers are generally constructed on the basis of wholesale prices because they are easily available.

(iv) Index numbers are generally based on samples. It is difficult and not possible to take into account each and every item in the calculation of index numbers. If the samples are not true representatives of the universe, index numbers calculated may not reflect the true picture of the problem.

(v) The index numbers may not reflect the relative change appropriately if the quality level of the items has changed.

(vi) Index numbers are liable to be misused if the base selected is not proper. The index numbers may give misleading results.

(vii) Index numbers constructed for one purpose cannot be used for another purpose. Every index number is constructed by a technique which is appropriate for the objective. For example, a wholesale price index number cannot measure cost of living. An index of production cannot measure productivity.

(viii) The major defect of index numbers is that they are based on past. Modern policies are framed on the basis of old data. These do not give correct results.

Preliminaries to the construction of the Index Numbers:

Before constructing the index numbers, certain preliminaries should be taken into consideration:

(i) The purpose of Construction must be defined as precisely as possible. The need for defining the purpose arises because the choice of method shall depend on the pourpose e.g. a cost of living index will have different set of commodities, price quotations and weights compared with a general wholesale price index.

(ii) Selection of base year: The general result is that the base year should be a normal period. The reference period should not be either too short or too long, it should be a normal or typical or current or representative period. This becaue people usually prefer to compare present conditions with conditions in the base period that is not too far back time. Base period should be free from all sorts of abnormalities and random or irregular fluctuations like earthquakes, wars, floods, famines, labour strikes, lock outs, economic boom and depression.

(iii) Selection of items: The prime consideration in the construc-

tion of the index numbers is the selection of commodities to be included and their prices. For example, to prepare the index number of the cost of living for higher income group, we must include only those items which are used by them e.g., the fuel items should be the electricity and gas cylinders and not the wood and coal.

(iv) Price quotations: Price quotations should be relevant to the purpose. For example, retail prices are more appropriate than wholesale prices for calculation of cost of living index. But retail prices differ from retailers to retailers and from customers to customers.

A choice has to be made from amongst them. The choice should be relevant to the purpose. There are two methods of quoting prices. (a) Money price and (b) Quantity Price. In case of money price, prices are quoted per unit of a comodity e.g., wheat at Rs.350 per quintal (100 Kilograms) while in case of quantity price-prices are quoted per unit of money e.g., wheat per rupee. Whatever be the mode of expression, It should be uniform for all the items and should preferably be in money price per unit of commodity.

It is not possible to obtain price quotations from all the places where a particular commodity is purchased or sold. Therefore, a selection of representative places and representative persons has to be made. Normally, such places are selected where the particular commodity is sold in large quantities. Generally representatives are appointed to collect the information. Care should be taken in selecting an agent, who should be unbiased.

(v) Number of commodities. As a rule, the larger the number of commodities, the more representative is the index numbers.But at the sametime, it must be remembered that if too many commodities are included, it will make the task of computation quite cumbersome. Hence, the number of items from which an index number is constructed should be fairly large consistent, with ease in handling them.

(vi) Choice of an average: For constructing an index number any method such as mean, median, mode, geometric mean and harmonic mean can be used. From the practical point of view arithmetic mean is generally used. Though with the development of the use of computors, the use of geometric mean is alsobecoming popular, because it renders the index number reversible.

(vii) To give weight: In order to allow each commodilty to have a reasonble influence on the index number, it is advisable to use a suitable weightage system. Unweighted number are those where all commodities are given equal importance. But in most cases, different commodities are given different degrees of importance,therefore, weights are assigned to

the various items. Allotment of weights is according to the purpose of the index number. Weighting may be done according to:

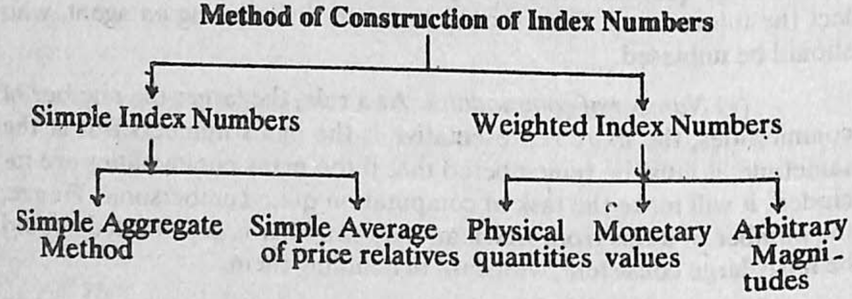
(a) Value or quantity produced

(b) value or quantity consumed, and

(c) value or quantity sold.

(viii) Choice of method: There are various methods of calculating index numbers such as aggregative method or the price relative method. Various methods have been proposed for calculation of weighted index numbers such as Laspeyre's method, Paasche's method, Dorbish and Bowley's methods, Marshall Edgeworth's method, Kelley's method and Fisher's method. *Fisher's method is considered as ideal for constructing index numbers.* No single formula can be appropriate for all types of index numbers and as such, the choice of a formula will have to be made taking into account the object of index numbers, the data available and the resources at the disposal of the person or organisation constructing the index number.

METHODS OF CONSTRUCTING INDEX NUMBERS



(A) Simple Index Numbers: They refer to those index numbers in which the commodities are of equal importance. There are two methods of constructing them:

(i) Simple Aggregate Method: Under this method of constructing an index number first, we obtain the sums of prices in the base year and current year. Then we divide the sum of prices of current year by the sum of prices in the base year and multiply the quotient by 100.

(ii) Simple Average of Price Relatives: Under this method first, we compare the price relative for different commodities. It is obtained

by dividing the price in the current year by the price in the base year and multiplying the quotient by 100. The average of the price relatives is then obtained by dividing the price in the current year by the price in the base year and multiplying the quotient by 100. The average of the price relatives is then obtained by using either the arithmetic mean or median or geometric mean. The resulting quotient gives us the results.

(B) Weighted Index Numbers: They refer to index numbers in which all comodities are given weight according to the importance of the commodity in a particular purpose. There are three types of weights which may be used in the construction of index numbers, but each type is applicable only under certain general methods of computations.

The system or types of weights are:

(a) Physical Quantities (b) Monetary Values, and (c) arbitrary Magnitudes.

The first type should be used only when the method of aggregate is used. The second type should be used only in weighting relatives. The third is an unscientific method and ordinarily should be used only in the crudest forms of index numbers.

SIMPLE INDEX NUMBERS

(i) Simple aggregate Method

Illustration No.1

Construct index numbers for 1988 taking 1984 as the base year from the following data by Simple Aggregative method:

Commodities:	A	B	C	D	E
Price in 1984(Rs):	100	80	160	220	40
Price in 1988(Rs):	140	120	180	240	40

Solution:

Steps:

(i) Add up the current year prices of various commodities.

(ii) Add up the base year prices of various commodities (Σp_0)

(iii) Use the following formula.

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

(p_{01} is price index of current year)

Commodities	Prices in 1984 (P_o)	Prices in 1988 (P_1)
A	100	140
B	80	120
C	160	180
D	220	240
E	40	40
	Σp_0 = 600	Σp_1=720

Applying formula, we get,

$$P_{o1} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{720}{600} \text{ x } 100$$

$$= 120$$

Thus the price index of 1988 is 120. It means that there is net increase in prices of commodities in the year 1988 to the extent of 20 per cent as compared to 1984.

Illustration No.2

Find the index numbers. (with the help of aggregate method) for 1988 by assuming 1985 as base year

Product	Unit per	Price 1985	Price 1988
Wheat	20 Kgs	60	70
Rice	20 Kgs	80	100
Pulses	20 Kgs	160	200
Ghee	Kg	10	12
Sugar	Kg	5	6

Solution:

Product	Unit per	Price 1985 (P_0)	Price 1988(P_1)
Wheat	20 Kgs	60	70
Rice	20 Kgs	80	100
Pulses	20 Kgs	160	200
Ghee	Kg	10	12
Sugar	Kg	5	6
		Σp_0 =315	Σp_1 =388

Simple aggregative Index Numbers:

$$p_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

$$= \frac{388}{315} \times 100$$

$$= 123.17$$

(ii) Simple Average of Price relatives Method

Illustration No.3

Construct Index number for 1987 taking 1985 as the base year from the following data by simple average of price relative method:

Commodities	A	B	C	D	E
Price in 1985(Rs):	100	80	160	220	40
Price in 1987(Rs):	140	120	180	240	40

Solution

Steps

(i) Calculate the price relatives of current year

$$\frac{p_1}{p_0} \times 100$$

(ii) Add up the calculated price relatives

$$\Sigma \frac{p_1}{p_0} \times 100$$

(iii) Divide the total of price relatives by number of commodities.

(iv) Apply the formula

$$p_{01} = \frac{\Sigma \frac{p_1}{p_0} \times 100}{N}$$

Index numbers by simple averages of price relative method:

Commodity	Price in 1985(p_0)	Prices in 1987(p_1)	Prices relatives $\frac{p_1}{p_0} \times 100$
A	100	140	$\frac{140}{100} \times 100 = 140$
B	80	120	$\frac{120}{80} \times 100 = 150$
C	160	180	$\frac{180}{160} \times 100 = 112.5$
D	220	240	$\frac{40}{220} \times 100 = 109.1$
E	40	40	$\frac{40}{40} \times 100 = 100$

$$\Sigma \frac{p}{p_0} \times 100 = 611.6$$

Now we get

$$p_{01} = \frac{\Sigma \frac{p}{p_0} \times 100}{N}$$

$$= \frac{611.6}{5} = 122.32 \text{ Ans.}$$

Fixed base and chain base methods of simple Index Numbers.

Illustration No.4

Prepare the price index number for the years 1984, 1985, 1986, 1987 and 1988 for the following data by considering 1984 as base with the help of (a) fixed method (b) chain method and (ii) 1988 as base with the help of (a) fixed method (b) chain method.

Commodities	Units	Price in rupees				
		1984	1985	1986	1987	1988
A	Per Kg	10	9	8	9	11
B	Per Kg	4	5	3	6	4
C	Per gm	6	8	5	4	6
D	Per litre	11	13	15	24	12
E	Per metre	6	8	10	14	12

Solution

Commodities	Unit	1984	1985	1986	1987	1988
A	Per Kg.	10	9	8	9	11
B	Per kg.	4	5	3	6	4
C	Per gm	6	8	5	4	6
D	Per litre	11	13	15	24	12
E	Per metre	6	8	10	14	12
		37	43	41	57	45

Index number with 1984 as base year:

Fixed Base:

Index Number for 1984 = 100

$$1985 = \frac{43}{37} \text{x} 100 = 116.2$$

$$1986 = \frac{41}{37} \text{x} 100 = 110.8$$

$$1987 = \frac{57}{37} \text{x } 100 = 154.0$$

$$1988 = \frac{45}{37} \text{x } 100 = 121.6$$

Chain Base:

Index Number for 1984 = 100

$$1985 = \frac{43}{37} \text{x } 100 = 116.2$$

$$1986 = \frac{41}{43} \text{x } 100 = 95.3$$

$$1987 = \frac{57}{41} \text{x } 100 = 139.0$$

$$1988 = \frac{45}{57} \text{x } 100 = 78.9$$

Index Number with 1988 as Base:

Fixed Base:

Index Number for 1988 = 100

$$1987 = \frac{57}{45} \text{x } 100 = 126.7$$

$$1986 = \frac{41}{45} \times 100 = 91.1$$

$$1985 = \frac{43}{45} \times 100 = 95.6$$

$$1984 = \frac{37}{45} \times 100 = 82.2$$

Chain Method:

Index Number for 1988 = 100

$$1987 = 100 \times \frac{57}{45} = 126.6$$

$$1986 = 100 \times \frac{41}{57} = 71.9$$

$$1985 = 100 \times \frac{43}{41} = 104.9$$

$$1984 = 100 \times \frac{37}{43} = 86.0$$

WEIGHTED INDEX NUMBERS

This method is also known as Aggregate Expenditure Method.

It is a very popular method. In this method appropriate weights are assigned to various commodities or items to reflect their relative importance in the group.

There are various methods of assigning the weights and for each method there is a different formula for construction of the index number:

Some of the important methods are:

(i) Laspeyre's Method

(ii) Paasche's Method

(iii) Fisher's Ideal Index Number

Formulae of Calculating Index Numbers:

1. Laspeyre's Method

$$p_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

The weights q_0 assigned to the commodities are weighted of the commodities of the base period.

2. Paasche's Method

$$p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

The weights q_1 assigned to the commodity(ies) are weights of the commodities of the current period.

(iii) Fisher's Ideal Index Number

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

Illustration No.5

Construct weighted index number on the basis of the following data with 1987 as the base year:

Commodities	1987		1988	
	Quantity	Price	Quantity	Price
A	8	10	10	9
B	15	15	12	30
C	7	5	14	6

Solution

Commodities	1987		1988		Products			
	Qty.	Price	Qty	Price				
	q_0	p_0	q_1	p_1	p_1q_0	p_1q_1	p_0q_1	$p_0\,q_0$
A	8	10	10	9	72	90	100	80
B	15	15	12	30	450	360	180	225
C	7	5	14	6	42	84	70	35
					564	534	350	340

Laspeyre's Method:

$$p_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0\, q_0} \times 100$$

$$= \frac{564}{340} \times 100 = 165.9 \text{ Ans.}$$

Paasche's Method:

$$p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

$$= \frac{534}{350} \times 100 = 152.6 \text{ Ans}$$

Fisher's Ideal Index Number

$$\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_1} \text{ x } \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \text{ x}100$$

$$= \sqrt{\frac{564}{340} \text{ x } \frac{534}{350}} \text{ x } 100 = \sqrt{1.66 \text{ x } 1.53 \text{ x }} 100$$

$$= \sqrt{2.52927 \text{ x}} 100$$

$$= 1.590 \text{ x } 100$$

$$= 159.0 \text{ Ans.}$$

Weighted Average of Price Relatives:

This mehod is an extension of the simple aggregate of relative method and is also known as Family Budget Method and in this method we make use of the value weights.

Illustration No.6

Calculate weighted average of price relative index number for 1988 on the basis of 1984 from the following data:

Commodities	Weight	Prices	Prices
	q_0	1984	1988
A	20	20	35
B	12	15	18
C	8	10	11
D	4	5	5
E	6	4	5

Solution:

Steps:

(1) Calculate the price relatives of the current year $(\frac{p_1}{p_0} \text{x}100)$, call it P

(2) Calculate value weights i.e., $p_0 q_0$ and term it as V

(3) Multiply P for each item by value weights i.e., $[(\frac{p_1}{p_0}\text{x}100) \times (p_0 q_0)]$

(4) Add the results obtained from multiplications carried out in step 3.

(5) Divide the total obtained by total value weights i.e., $\Sigma p_0 q_0$.

(6) Use following formula:

$$p_{01} = \frac{\Sigma \left[\left(\frac{p_1}{p_0} \text{ x } 100 \right) \text{ x } (p_0 q_0) \right]}{\Sigma p_0 q_0}$$

Commodities	Weights (W)	Price 1984	Price 1988		Price Relatives	
	q_0	p_0	p_1	(V) ($p_0 q_0$)	P	PV
A	20	20	35	400	175	70000
B	12	15	18	180	120	21600
C	8	10	11	80	110	8800
D	4	5	5	20	100	2000
E	6	4	5	24	125	3000
	ΣW = 50			ΣV = 704	ΣP x V = 105400	

In the above case, we get:

$$p_{01} = \frac{\Sigma PV}{\Sigma V} = \frac{105400}{704} = 149.71$$

CONSUMER PRICE INDEX NUMBERS

These index numbers are also called (i) Cost of Living Index Numbers, (ii) Retail Price Index Numbers.

These Index Numbers tell us how much the consumers of a particular class have to pay to get a basket of goods and services at a particular point of time in comparision to what they paid for this basket in the base year.

These index numbers are constructed separately for different classes or people or groups or sections of the society, *e.g.*, urban wage earners, agricultural labourers, industrial workers, government employees etc. and also for different geographical areas like town, city, rural area, hilly area, *etc.*

These index numbers are very useful in life. Following are some of their advantages:

(i) Trade Union leaders can make wage negotiations on the basis of consumer price index numbers.

(ii) They are used to measure the purchasing power of rupee and thereby helping to find out the real income of a particular group.

(iii) They provide guidelines to the government for deciding wage policy, control of prices and general economic policies.

(iv) Such index numbers are used in analysing the markets for particular kinds of goods and serives.

METHODS OF CONSTRUCTION OF CONSUMER INDEX NUMBERS

There are two methods of constructing consumer price index numbers:

(i) Aggregate Expenditure Method or Aggregative Method.

$$\text{Consumer Price Index} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

This is the most popular method.

(ii) Family Budget or Weighted Relatives:

Consumer Price Index:

$$= \frac{\Sigma PV}{\Sigma V} \text{ or } \left[\frac{\Sigma IW}{\Sigma W}\right]$$

Here $P = \frac{p_1}{p_0} \times 100$ for each item

V = Value weights $(p_0\ q_0)$

This method is the same as the weighted average of price relatives.

Illustration No.7

Construct the consumer price index number for 1987 on the basis of 1980 from the following data by using

(i) Aggregative Method

(ii) Weighted Relatives Method.

Commodity	Rice	Wheat	Pulses	Ghee	Oil
Weights (q_0)	40	25	5	20	10
Price 1980	16	40	0.50	5.12	2.00
Price 1987	20.0	60.0	0.50	6.25	1.50

Solution

Aggregative Expenditure Method

Commodity	Weights q_0	Price 1980 p_0	Price 1987 p_1	$p_0 q_0$	$p_1 q_0$
Rice	40	16.00	20.00	640.00	800.00
Wheat	25	40.00	60.00	1000.00	1500.00
Pulses	5	0.50	0.50	2.50	2.50
Ghee	20	5.12	6.25	102.40	125.00
Oil	10	2.00	1.50	20.00	15.00
				1764.90	2442.50

Consumer Price Index for 1987.

$$= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{2442.50}{1764.90} \times 100 = 138.39$$

Weighted Relatives Method OR Family Budget Method

Commodity	Weights W	Price 1980 p_0	Price 1987 p_1	Price Relatives $\frac{p_1}{p_0} \times 100$ (I)	Weighted Relative IW
Rice	40	16.00	20.00	125	5000
Wheat	25	40.00	60.00	150	3750
Pulses	5	0.50	0.50	100	500
Ghee	20	5.12	6.25	122	2440
Oil	10	2.00	1.50	75	750
	$\Sigma W = 100$				$\Sigma IW = 12440$

$$\text{Cost of Living Index} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{12440}{100}$$

$= 124.4$ Ans.

QUESTIONS

1. From the following data prepare a price index number with 1981 as the base year:

Year	Price per Quintal
1981	125
1982	150
1983	140
1984	160
1985	175

2. State the uses of index numbers. What considerations are taken into account while selecting the base year for the preparation of the index number?

(Hint: Base year selection needed-normal year, should not be too far away from the year of comparison.)

3. Write short notes on:

(a) Weighted Index Numbers.

(b) Consumer Price Index Numbers.

(c) Use of Index Numbers.

(d) Index Numbers.

4. Explain the need and problem of "weightage in the construction of index numbers".

(Hint: There are number of commodities included in the construction of an index number. All the commodities are not of equal significance in our daily life. So we should attach some weight to the commodities on the basis of the significance which they have in our daily life).

5. Define index numbers. Using the Simple Aggregate Method, calculate the price index number for 1988 (Base 1980) from the following data:

Commodities:	A	B	C	D	E
1980 Prices(Rs):	50	40	10	5	2
1988 Prices(Rs):	80	60	20	10	6

6. Define index numbers. What are the problems in the construction of index numbers? Compute index numbers for the year 1971 to 1975 from the following data:(1970 base year). Use Price Relative Method.

Year:	1970	1971	1972	1973	1974	1975
Price:	10	14	16	20	22	24

7. Briefly explain the importance of index numbers in the study of economics. Using the Simple Aggregative Method, calculate the price index for 1979 (base year 1975) from the following data:

Commodities	A	B	C	D	E
1975Prices(Rs):	55	40	10	5	2
1979 Prices(Rs):	80	60	20	10	6

8."Index Numbers are economic barometers". Explain this statement.

9. Calculate Price Index Number for 1988 from the following data by Weighted Aggregative Method using:

a) Laspeyre's Method

b) Paasche's Method

c) Fisher's Ideal Index Number.

Commodity	Price 1975	Quantity 1975	Price 1984	Quantity 1984
A	4	20	6	10
B	3	15	5	23
C	2	25	3	15
D	5	10	4	40

10. Construct cost of living index number:

	Price Relatives	Weights
Food	250	45
Rent	150	15
Clothing	320	20
Fuel and Lighting	190	5
Miscellaneous	300	15

11. Construct the cost of living index for 1987 from the following data:

Group	Weights	Price Indices
Food	42	247
Fuel and lighting	7	293
Clothing	8	289
House Rent	13	100
Miscellaneous	14	236

12. Figures from the family budget of some families are given for 1987 and 1988. What changes in the cost of living figures for 1988 as compared with that of 1987 are seen?

Expenses on	Food	Rent	Clothing	Fuel	Others
Percentages	30%	15%	20%	10%	25%
Prices in 1987(Rs)	100	20	70	20	40
Prices in 1988 (Rs)	90	20	60	15	55

13. From fixed base index number given below, prepare chain base index numbers:

Years	1975	1976	1977	1978	1979	1980
Indices	94	95	102	95	98	100

14. From the cahin base index number, given below, prerpare fixed base index numbers:

Years	1980	1981	1982	1983	1984
Indices	89	110	120	90	40

15. The following table gives the prices of some commodities in the base year and the quantity sold in the base year. Calculate the weighted index number by using the weighted average of relatives method:

Commodities	Base year Quantity (in units)	Base year Price (in Rs)	Current year Price (in Rs)
A	7	16	18.6
B	6	2	3.2
C	15	5.6	7.0
D	21	1.5	1.4

16. From the data given below construct an Index number for 1988 taking 1981 as base year by the average relatives Method:

Items	1981 Price (in Rs)	1988 Price (in Rs)
A	60	75
B	45	60
C	80	96
D	110	121
E	15	109

17 .The price quotations of different commodities for 1984 and 1988 are given below. Calculate the index numbers for 1988 with 1984 as base being: (i) the simple average of price relatives, (ii) the weighted average of price relatives.

(price in rupees)

Commodity	Unit	Weights	1984	1988
A	Kg	5	2.00	4.50
B	Quintal	7	2.50	3.20
C	Dozen	6	3.00	4.50
D	Litre	2	1.09	1.80

APPENDIX

List of Symbols and Formulae

SYMBOLS

x = denote the variable or the mid-value of the class interval.

N = total number of given observations or items in a group or series.

Σ = Sum of the values.

f = frequencies

Σx = sum of the values of the variables

Σf = sum of the frequencies

A = asumed or arbitrary mean

i = width of the class interval

d = deviation of any single value of the variable

d' = step deviation

M = Median

Z = Mode

l_1 = lower limit of the group

Q_1 = First quartile

Q_3 = Third quartile

D = Decile

P = Percentile

f_1 = frequency of the class preceeding the model class

f_2 = frequency of the class succeeding the model class

fm = frequency of the model class

dm = absolute deviation of the value of the variable from the median.

dx = absolute deviation of the value from the mean.

M.D. = Mean Deviation

σ = Standard deviation

r = Co-eficient of Corelation due to Pearson

R = Coefficient of rank correlation due to Spearman

p_0 = price of base year

q_0 = quantity of base year

p_1 = price of current year

q_1 = quantity of current year

w = weights

FORMULAE

Arithmetic Mean

(a) Individual series

$$\overline{X} = \frac{\Sigma x}{N}$$

(i) Direct Method

Here Σx = sum of values of variable

N = sum of items.

(ii) Assumed Mean Method

$$\overline{X} = A \pm \frac{\Sigma dx}{N}$$

Here A is assumed Mean

Σdx is sum of deviations taken from assumed mean.

(b) Discrete Series

(i) Direct Method

$$\overline{X} = \frac{\Sigma fx}{N}$$

Here Σfx is total of variable multiplied by their frequencies.

(ii) Short cut Method

$$\overline{X} = A \pm \frac{\Sigma fdx}{N}$$

Here Σfdx is sum total of deviations from A multiplied by frequencies.

(iii) Step Deviation Method

$$\overline{X} = A \pm \frac{\Sigma fdx'}{f} \text{ x i}$$

Here Σfdx' is taken from dx after taking out a common value and is multiplied by f (frequencies) "i" is the common factor.

(c) Class Interval Series

(i) Direct Method $\overline{X} = \frac{\Sigma fx}{f}$

(ii) Shortcut Method

$$\overline{X} = A \pm \frac{\Sigma fdx}{f}$$

(iii) Step deviation Method

$$\overline{X} = A \pm \frac{\Sigma fdx'}{f} \text{ x i}$$

Weighted Arithmetic Mean

Weighted mean $= \overline{X}W = \frac{\Sigma wx}{W}$

Here W refers to weights

MEDIAN

(A) Individual Series

$$M = \text{value of } \frac{N+1}{2} \text{ the item}$$

(B) Class Intervals

$$M = l_1 + \frac{\frac{N}{2} - C}{f} \text{ x i}$$

Here "l_1" is the lower limit of the median group.

"i" is the width of the class interval, where the median lies.

N = is the total number of frequencies

f = is the frequency of the median class.

MODE

Class-interval Series

$$= l_1 + \frac{f_m - f_1}{2fm - f_1 - f_2} \times i$$

Here "l_1" is the lower limit of the model class.

"fm" is the frequency of the model group.

"f_1" is the frequency of premodel group.

"f_2" is the frequency of the post-model group.

"i" is the width of the class-interval.

Quartiles

$$Q_1 = l_1 + \frac{\frac{N}{4} - C}{f} \times i$$

$$Q3 = l_1 + \frac{\frac{3XN}{4} - C}{f} \times i$$

$$= l_1 + \frac{\frac{3N}{4} - C}{f} \times i$$

Here "l_1" is the lower limit of class interval in which Q_1 lies. "C" is the cumulative frequency of the class interval preceeding the quartile class

"i" is the width of class interval.

"f" is the frequency of the class in which quartile lies.

Deciles

$$D1 = l_1 + \frac{\frac{N \times j}{10} - C}{f} \times i$$

Here "j" is the decile required.

Percentiles

$$P_1 = l_1 + \frac{\frac{N \times k}{100} - C}{f} \times i$$

Here "k" is the required number of Percentile.

Measurement of Variability

Range = L-S

Here "L" stands for largest value

and "S" stands for smallest value

Co-efficient of Range $= \frac{L-S}{L+S}$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

Inter Quartile range $= Q_3 - Q_1$

Co-efficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

MEAN DEVIATION (Individual Series)

$$M.D. = \frac{\Sigma\,|dx|}{N}$$

Co-efficient of Mean deviation $= \frac{M.D.}{Mean}$

Median Deviation

$$M.D. = \frac{\Sigma|\,dm\,|}{N}$$

Coefficient of Median Deviation $= \frac{M.D.}{Median}$

Mean Deviation (Discrete Series)

Mean Deviation $= \frac{\Sigma f\,|dx|}{N}$

Median Deviation $= \frac{\Sigma f\,|dm|}{N}$

Standard Deviation

$$= \sqrt{\frac{\Sigma dx^2}{N}}$$

$$= \sqrt{\frac{\Sigma f\,(dx)^2}{N}}$$

$$= \sqrt{\left(\frac{\Sigma fdx^2}{\Sigma f}\right) - \left(\frac{\Sigma fdx}{\Sigma f}\right)} \times i$$

Coefficient of Variation

$$\text{C.V.} = 100 \times \frac{\text{S.D.}}{\text{Mean}}$$

KARL PEARSON'S CORRELATION COEFFICIENT

$$r = \frac{xy}{N \times \sigma x \times \sigma y}$$

Where $x = (x - \bar{x})$

$$y = (y - \bar{y})$$

$$r = \frac{\Sigma xy}{N \times \sqrt{\frac{\Sigma x^2}{N}} \times \sqrt{\frac{\Sigma y^2}{N}}}$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx \times \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

$$r = \frac{\Sigma xy - \frac{1}{N} \Sigma x \Sigma y}{\sqrt{\Sigma x^2 \frac{(\Sigma x)^2}{N}} \times \sqrt{(\Sigma y)^2 - \frac{(\Sigma y)^2}{N}}}$$

Spearman's Rank Correlation

$$R = 1 - \frac{\sigma \Sigma D^2}{N(N^2 - 1)}$$

INDEX NUMBER

Simple Aggregate Method

$$p_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

Simple Average of price relatives method

$$p_{01} = \frac{\Sigma \frac{p_1}{p_0} \times 100}{N}$$

Weighted Index Numbers

Laspeyre's Method

$$p_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Pasche's Method

$$p_{01} = \frac{\Sigma p_1\, q_0}{\Sigma p_0 q_1} \times 100$$

Fisher's Ideal Index Number

$$p_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0\, q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

Weighted Average of Price Relatives

$$p_{01} = \frac{\Sigma \;\left(\frac{p_1}{p_0} \times 100\right) \times \Sigma\,(p_0 q_0\;)}{\Sigma p_0\, q_0}$$

$$p_0 q_1 = \frac{\Sigma PV}{\Sigma V}$$

When Base year value weights are used

Consumer Price Index Numbers

(a) Aggregate Expenditure Method

$$\text{Consumer Index Number} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

(b) Family Budget or Weighted relatives method

$$\text{Consumer Price Index} = \frac{\Sigma PV}{\Sigma V}$$

$$\text{or} \qquad = \frac{[\Sigma IW]}{\Sigma W}$$

$$\text{Here } P = \frac{p_1}{p_0} \times 100 \text{ for each item.}$$

$$V = \text{value weights } (p_0 q_0)$$

ANSWERS

Unit IV

Chapter 1

Mean

1. 14.625
3. 28.36
4. 38.8
5. 35.5
6. 26.5
7. 45.00
8. 51.00 approx
9. 37.5
10. 22.8
11. 57.4
12. 834.61
14. 28.6
15. 21.9
16. 5.00
17. 773.00
18. 51.67
19. 153.64
21. 50.4
22. 28.6
23. 30.08

Unit IV

Chapter 2

Median

2. 143.3
3. 30.78
4. 61.00

5. 37.00
6. 22.9
7. 22.00
8. 59.35
9. 65.6
10. 36.3
11. 52.7
12. 61.00
13. 5.00
14. 153.79
15. 40.00
16. 28.3
17. 59.35
18. 45.43
19. 250.00
20. 43.125 years

Unit IV

Chapter 3

Mode

2. 354.54
3. 38
3 . 66.8
5 . 4.0
6 . 57.5
7. 51.09
8. 31.5
9. 23.46
10. 46.15
11. 35.6

Unit IV

Chapter 4

Quartiles, Deciles and Percentiles

1. Med 61, $Q_3 = 63$
2. Med 22.9, $Q_3 = 40.8$, $P_{70} = 37.14$
3. Med 59.35, $Q_3 = 68.52$, $Q_1 = 31.04$
4. Med 61.
5. Med 28.3, $Q_1 = 18.75$, $Q_3 = 37.7$
6. Med 40, $Q_1 = 34$, $Q_3 = 47.08$
7. Med 45.43, $Q_1 = 38.4$, $Q_3 = 51.1$
8. $D_6 = 50$, $P_{78} = 58.18$ $P_{65} = 52.27$
9. $Q_1 = 15.85$, $Q_3 = 36.5$, $D_7 = 34.08$, $P_{80} = 50.2$
10. $Q_1 = 22.9$, $Q_3 = 38$
11. $Q_1 = 34.16$, $Q_3 = 57.5$, $D_6 = 50$, $P_{35} = 41.07$

Unit IV

Chapter 5

3. 2.9
4. Mode $= 46.15$, S.D. $= 17.26$
5. Mean $= 38.8$, M.D. $= 17.06$,
   Coefficient of M.D. $= 0.439$
7. Mean $= 45$, S.D. $= 15.9$
8. Mean 51 approx., M.D. $= 14.4$ approx
10. Mean $= 22.8$, S.D. $= 11.83$
11. S.D. $= 788.4$
13. S.D. $= 9.1$
15. Mean $= 28.6$, S.D. $= 12.2$
17. (a) C.V. of A school 15, B school 15, Variability is equal in both the cases.

(b) Variance $= 8.66$

18. Mean $= 21.9$, S.D. $= 7.99$
19. (c) Mean $= 773$, M.D. $= 64.4$

20. Mean = 51.67, S.D. = 15.13
21. S.D. = 12.33
22. S.D. = 8.19
23. Mean = 6, M.D. = 2.13, S.D = 3.26

Unit IV

Chapter 6

Measures of Correlation

1. 0.6
2. 0.093
3. 0.91
4. 0.745
6. 0.57
8. 0.18
9. 0.96
10. 1

Unit IV

Chapter 7

Index Number

1. 1981 = 100, 1982 = 120, 1983 = 112, 1984 = 128, 1985 = 140
5. Index for 1982 = 164.48
6. Price relative Index Number = 176.66
7. Index for 1979 = 157.14
9. (Laspeyre's-137.77, Paasche's-158.99, Fisher's = 148)
10. Cost of living Index Number = 253.5
11. Cost of living Index Number = 231.19
12. Index Number for 1988 = 97, a decrease of 3% in the cost of living.
13. 100, 104.25, 104.1, 93.14, 103.16, 102.04.
14. 80, 88, 105.6, 95.04, 38.02
15. 121.5
16. Price Index Number 108.0
17. Simple Average of Price Relatives= 170.75

    Weighted Average of Price Relatives= 164.05

S. C. E. R. T. West
Library
Calcutta
Road